The Potential Regulatory Role of microRNA in Methamphetamine Use Disorder (MUD)

by

David Christopher Newey

Dissertation presented for the degree of Master of Science in the Faculty of Science

at

Stellenbosch University

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged (grant holders: Christine Lochner and Nathaniel McGregor). Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Supervisor: Dr Nathaniel Wade McGregor

Co-Supervisors: Prof Christine Lochner, Dr Kevin Sean O'Connell

December 2020

Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 07/2020

Abstract

Methamphetamine (MA) is a psychostimulant affecting the central nervous system. Individuals with methamphetamine use disorder (MUD) present compulsive substance-seeking behaviour and impaired control over use. Use of MA is associated with symptoms such as euphoria, hyperalertness and impairments in executive function and working memory. MA is a common substance of abuse in South Africa. MUD is a multifactorial psychiatric condition with its aetiology involving a complex interplay between genes and environmental factors. Evidence suggests that mechanisms regulating behavioural abnormalities associated with MUD involve changes in gene expression throughout the brain's reward circuitry. Genome-wide association studies (GWAS) have revealed genes that may be involved in MUD. Additionally, a role has been established for epigenetic mechanisms, such as miRNA, mediating the effects of MA on the brain. There is an increase in studies investigating miRNA involvement in MUD. Considering links between environment, miRNA and neuropsychiatric disorders and the overlap in the molecular pathways of these disorders, it is likely that these pathways are regulated differently resulting in their differing clinical manifestations. Hence, this study aimed to elucidate the role of miRNA-mediated regulation in MUD and further disentangle the molecular underpinnings of MUD in a South African context. This was accomplished through a regression analysis of data from a local cohort with a diagnosis of MUD before performing *in silico* analyses on MUD data and a discovery cohort with cocaine use disorder (CUD); as both MA and cocaine are classed as psychostimulants. The MUD cohort was genotyped and imputed before being used in a regression analysis to identify single-nucleotide polymorphisms (SNPs) associated with MUD. Principle component analysis (PCA) was performed to investigate the effects of population stratification on the outcome of this analysis. Subsequently, associated SNPs from the MUD and CUD cohorts were investigated using *in silico* analyses to determine host genes for associated SNPs. These genes were compared to identify those exclusive to the MUD cohort, and were subsequently enriched to identify associated biological pathways and miRNA.

The regression analysis identified 510 SNPs approaching significant association ($p<1x10^{-4}$) with MUD. The genes identified in the MUD and CUD cohorts were compared, which led to the identification of 57 genes exclusively associated with the MUD cohort. These genes were found to be associated with several pathways involved in the aetiology of MUD such as autophagy and apoptosis. The genes were also regulated by miRNA previously associated with MUD.

In conclusion, this study was able to identify several miRNA and genes trending towards significance. These findings are consistent with the current literature on MUD and contribute to knowledge on the molecular underpinnings of MUD by highlighting differences between MUD and other stimulant use disorders. The findings identify an epigenetic component to MUD aetiology via miRNA and speak

to underlying regulatory networks involved in MUD aetiology. This is the first study investigating the molecular underpinnings of MUD in a South African cohort, indicating the potential for use of local populations to identify novel variants associated with miRNA-mediated regulation in MUD aetiology.

Opsomming

Methamphetamine (MA) is 'n psigostimulant wat die sentrale senuweestelsel beïnvloed. Persone met metamfetamien-gebruiksversteuring (MUD) bied kompulsiewe gedrag op soek na middels en verswakte beheer oor gebruik. Die gebruik van MA word geassosieer met simptome soos euforie, hiperalertiteit en gestremdhede in die uitvoerende funksie en werkgeheue. MA is 'n algemene misbruik in Suid-Afrika. MUD is 'n multifaktoriale psigiatriese toestand met sy etiologie wat 'n komplekse wisselwerking tussen gene en omgewingsfaktore behels. Getuienis dui daarop dat meganismes wat gedrags abnormaliteite reguleer wat verband hou met die spysverteringskanaal, veranderings in geenuitdrukking deur die hele brein se beloningskringloop behels. Genoomwye assosiasiestudies (GWAS) het gene geopenbaar wat by MUD betrokke kan wees. Daarbenewens is 'n rol gevestig vir epigenetiese meganismes, soos miRNA, wat die effekte van MA op die brein bemiddel. Daar is 'n toename in studies wat miRNA-betrokkenheid by MUD ondersoek. Met inagneming van skakels tussen omgewings-, miRNA- en neuropsigiatriese afwykings en die oorvleueling in die molekulêre weë van hierdie afwykings, is dit waarskynlik dat hierdie weë anders gereguleer word, wat lei tot hul verskillende kliniese manifestasies. Daarom het hierdie studie ten doel gehad om die rol van miRNA-bemiddelde regulering in MUD toe te lig en die molekulêre onderbou van MUD in 'n Suid-Afrikaanse konteks verder te ontwrig. Dit is bewerkstellig deur 'n regressie-analise van data van 'n plaaslike kohort met 'n diagnose van MUD voordat dit in silico-analises op MUD-data uitgevoer is en 'n ontdekkingskohort met kokaïengebruiksteuring (CUD); aangesien beide MA en kokaïen as psigostimulante geklassifiseer word. Die MUD-kohort is genotipeer en toegereken voordat dit in 'n regressie-analise gebruik is om enkel-nukleotied polimorfismes (SNP's) geassosieer met MUD te identifiseer. Beginselkomponentanalise (PCA) is uitgevoer om die gevolge van populasie-stratifikasie op die uitkoms van hierdie analise te ondersoek. Gevolglik is geassosieerde SNP's van die MUD- en CUD-kohorte ondersoek in silico-analises om gasheergenes vir geassosieerde SNP's te bepaal. Hierdie gene is vergelyk met die identifisering van dié wat eksklusief tot die MUD-kohort was, en is vervolgens verryk om gepaardgaande biologiese weë en miRNA te identifiseer.

Die regressie-analise het 510 SNP's geïdentifiseer wat beduidende assosiasie ($p < 1x10^{-4}$) met MUD nader. Die gene wat in die MUD- en CUD-kohorte geïdentifiseer is, is vergelyk, wat gelei het tot die identifikasie van 57 gene wat eksklusief met die MUD-kohort verband hou. Daar is gevind dat hierdie gene geassosieer word met verskillende weë wat betrokke was by die etiologie van MUD, soos outofagie en apoptose. Die gene is ook gereguleer deur miRNA wat voorheen met MUD geassosieer is.

Ten slotte kon hierdie studie verskillende miRNA en gene identifiseer wat na betekenisvolheid was. Hierdie bevindinge strook met die huidige literatuur oor MUD en dra by tot kennis oor die molekulêre onderbou van MUD deur die verskille tussen MUD en ander stimulantegebruiksversteurings uit te lig. Die bevindinge identifiseer 'n epigenetiese komponent vir MUD-etiologie via miRNA en spreek tot onderliggende regulatoriese netwerke wat betrokke is by MUD-etiologie. Dit is die eerste studie wat die molekulêre onderbou van MUD in 'n Suid-Afrikaanse kohort ondersoek het, wat 'n aanduiding is van die potensiaal vir die gebruik van plaaslike populasies om nuwe variante te identifiseer wat verband hou met miRNA-bemiddelde regulering in MUD-etiologie.

Acknowledgements

I wish to express my sincere gratitude and appreciation to the following persons and institutions:

- Dr Nathaniel Wade McGregor for his guidance throughout this study and his unwavering support
- Prof Christine Lochner for her constant assistance in furthering my understanding of the clinical aspects of the study
- Dr Kevin Sean O'Connell for his assistance in improving the work done to a more professional level
- The National Research Foundation of South Africa for their financial support
- Stellenbosch University for creating a space where this research can be performed in an adequate capacity
- Lab 231 (Ms Emma Frickel, Ms Megan Hamilton, Ms Ellen Ovenden, Mr Wilro van Niekerk and Mr Wade Ambrose) for their support throughout the study both technically and morally

Table of Contents

List of Acronyms

ASD – Autism Spectrum Disorder

BD – Bipolar Disorder

CUD – Cocaine Use Disorder

DNA – Deoxyribonucleic Acid

DSM-IV – Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition

DSM-5 – Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition

GAD – Genetic Association Database

GO – Gene Ontology

GWAS – Genome-wide Association Study

HWE – Hardy-Weinberg Equilibrium

MA - Methamphetamine

MAF – Minor Allele Frequency

MDD – Major Depressive Disorder

mRNA – Messenger RNA

miRNA – Micro RNA

miR – Micro RNA (prefix)

MUD – Methamphetamine Use Disorder

NP – Neuropsychiatric

nt – Nucleotide

PCA – Principle Component Analysis

PGC – Psychiatric Genomics Consortium

SNP – Single Nucleotide Polymorphism

SUD – Substance Use Disorder

SZ - Schizophrenia

UTR – Untranslated Region

List of Figures

Figure 6 - Manhattan plot of $-\log_{10}(P)$ logistic regression results for MUD cohort SNPs across each chromosome. A blue 'trend towards significance line' indicates SNPs passing the nominal significance threshold of $1x10^{-4}$ (n = 510).

Figure 7 - Q-Q plot of $-\log_{10}(P)$ results of logistic regression analysis of surviving MUD cohort SNPs plotted against their expected values. Lambda (λ) (genomic inflation) was calculated to be 2.08 by Plink.

Figure 8 - Manhattan plot of –log10(P) logistic regression results for MUD cohort SNPs across each chromosome after adjusting for population stratification using PCs. As expected, once population stratification is adjusted for, the power to detect association decreases and no SNPs pass the nominal (1x10-4) or genome-wide (5x10-8) significance thresholds.

Figure 9 - Q-Q plot of –log10(P) results of logistic regression analysis of surviving MUD cohort SNPs, after adjusting for population stratification using PCs, plotted against their expected values. Lambda (λ) (genomic inflation) was calculated to be 1.20 by Plink.

List of Tables

List of Supplementary Tables and Figures

B) Histograms visualising the heterozygosity-based distributions measured either using F statistics (calculated by plink) or H (calculated using observed homozygotes and the number of non-missing autosomal genotypes).

C) Quantile plots of genotype call rates and SNP coverage were used to confirm the viability of the data for further analyses. These indicate the SNPs passing the threshold of 95% (i.e. 95% of individuals in the study were not missing data for aforementioned SNPs), and the coverage of the SNPs read-depths in proportion to the quantiles of the theoretical cumulative distribution function with SNPs presenting low coverage being removed from the data after quality control.

Supplementary Figure 2 - Manhattan plot of $-\log_{10}(P)$ logistic regression results for MUD cohort SNPs across each chromosome (where chr 24 refers to the sex chromosomes). A blue 'trend towards significance line' indicates SNPs passing the nominal significance threshold of $1x10^{-4}$ (n = 125).

Supplementary Figure 3 - Q-Q plot of $-\log_{10}(P)$ results of logistic regression analysis of surviving MUD cohort SNPs plotted against their expected values. Lambda (λ) (genomic inflation) was calculated to be 1.57485 by Plink.

Supplementary Figure 4 - Manhattan plot of –log10(P) logistic regression results for MUD cohort SNPs across each chromosome (where chr 24 refers to the sex chromosomes) after adjusting for population stratification using PCs. As expected, once population stratification is adjusted for, the power to detect association decreases and no SNPs pass the nominal (1x10-4) or genome-wide (5x10-8) significance thresholds.

Supplementary Figure 5 - Q-Q plot of –log10(P) results of logistic regression analysis of surviving MUD cohort SNPs, after adjusting for population stratification using PCs, plotted against their expected values. Lambda (λ) (genomic inflation) was calculated to be 1.15717 by Plink.

Chapter 1 Literature Review

1.1 Introduction

With the increasing awareness of mental disorders it is not surprising to find that there is currently an increase in the incidence of neuropsychiatric (NP) disorders (Baxter *et al.*, 2013). Studies have indicated that groups of individuals in the lower tiers of social hierarchies tend to suffer harsher psychological stressors; this is of paramount importance in countries like South Africa where there is a history of political and social marginalisation and oppression (Lanesman *et al.*, 2019). This adversity is associated with an increase in vulnerability to mental disorders like substance use disorders (SUD).

The 5th edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) defines SUD as a condition spanning a wide variety of problems arising from taking the substance in larger amounts or for longer than was meant, loss of control over use (i.e. wanting to cut down or stop using the substance but not managing to, and development of withdrawal symptoms, which can be relieved by taking more of the substance), spending a lot of time getting, using, or recovering from use of the substance, cravings and urges to use the substance and continuing to use despite impairment in social, occupational, or recreational activities (APA, 2013; Prom-Wormley *et al.*, 2017). Individuals using methamphetamine (MA) experience multiple symptoms such as euphoria, increased alertness, lack of appetite, deficits in episodic memory and impairments of executive function (Farhadian *et al.*, 2017). The levels of impairment to memory and executive function may depend on the amount of time an individual has been abusing MA and often lead to difficulties during treatment due to the problems caused by episodic memory and executive function impairments (Farhadian *et al.*, 2017). MUD has been associated with impairments to social, work and family functionality at interpersonal, intrapersonal and community levels (Sommers, Baskin and Baskin-Sommers, 2006; Watt *et al.*, 2015)

Substance abuse is associated with severe socioeconomic and public health problems particularly in South Africa where a study found that 4.4% (n = 26453) of individuals had used illicit substances with 0.8% of individuals having used MA, within the last 3 months (Peltzer and Phaswana-Mafuya, 2018). Of the substances abused locally, MA is one of the more commonly abused; this is especially evident in the Western Cape where 33% of individuals seeking specialist substance abuse treatment is due to either MA-induced psychosis or methamphetamine use disorder (MUD) (Thomas *et al.*, 2016). Methamphetamine is a highly addictive psychostimulant affecting various monoamine neurotransmitter systems, such as dopamine and serotonin, and results in feelings of alertness, increasing energy, and euphoria (Yu *et al.*, 2015). In DSM-5 MUD is classified as a stimulant use disorder along with cocaine use disorder (CUD) (American Psychiatric Association, 2013). This is

due to the overarching clinical similarities between stimulant class drug effects (American Psychiatric Association, 2013; Jensen, 2016; Substance Abuse and Mental Health Services Administration, 2016). The DSM diagnostic criteria for MUD have changed over the several editions of the DSM, with several minimum criteria required for diagnosis changing across editions. For example, the DSM-5 combined the previously separate MUD and CUD into a single category (stimulant use disorder) (Substance Abuse and Mental Health Services Administration, 2016). The overall changes between the DSM-IV and DSM-5 relating to SUDs can be demonstrated by Table 1 below (Substance Abuse and Mental Health Services Administration, 2016).

Table 1: Modified version of Table 2.1 from Substance Abuse and Mental Health Services Administration (2016) detailing the differences relating to SUDs between the DSM-IV and DSM-5; notably several criteria relating to diagnosis underwent changes and MUD and CUD were combined under stimulant use disorder.

Characteristic	DSM-IV	DSM-5
Disorder Class	Substance-related disorders, included only SUDs	Substance-related and addictive disorders class now includes SUDs and gambling disorder (formerly pathological gambling)
Disorder Types	Abuse and dependence hierarchical diagnostic rules meant that people ever meeting criteria for dependence did not receive a diagnosis of abuse for the same class of substance	SUD, substance abuse and dependence have been eliminated in favour of a single diagnosis, SUD
Substances Assessed	11 classes of substances assessed, plus 2 additional categories	10 classes of substances assessed, plus 2 additional categories
	• Alcohol	• Alcohol
	• Amphetamine and similar sympathomimetics	• Stimulant use disorder, which includes amphetamines (including methamphetamine), cocaine, and other stimulants

	• Caffeine (intoxication only)	• Caffeine (intoxication and withdrawal)
	• Cannabis (no withdrawal syndrome)	• Cannabis (with withdrawal syndrome)
	• Cocaine	• Combined with other stimulants (e.g., amphetamines) under stimulant use disorder
	• Hallucinogens • Phencyclidine and similar arylcyclohexylamines	• Separated into phencyclidine use disorder and other hallucinogen use disorder
	• Inhalants (no withdrawal syndrome)	• Inhalants (no withdrawal syndrome)
	• Nicotine (dependence only)	• Tobacco
	• Opioids	• Opioids
		• Merged with hallucinogens
	• Sedatives, hypnotics, and anxiolytics	• Sedatives, hypnotics, and anxiolytics
	• Other drug abuse/dependence	• Any other SUD
	• Polysubstance dependence	• Dropped polysubstance use disorder
Disorders Assessed	Substance abuse: One or more symptoms Substance dependence: Three or more symptoms in the same 12-month period (or one symptom if dependence criteria have been met previously in the lifetime)	SUD: Two out of 11 criteria clustering in a 12-month period are needed to meet disorder threshold
Severity	No severity criteria	Severity is assessed in terms of the number of symptoms that meet criteria:

		• Mild: two to three symptoms • Moderate: four to five symptoms • Severe: six or more symptoms
Additional Specifications	With or without physiological dependence, early full remission, early partial remission, sustained full remission, sustained partial remission, on agonist therapy, and in a controlled environment	-

The clinical effects of MA use vary depending on length of use, method of usage and dosage used (Ciccarone, 2011). The higher the dosage used by individuals the greater the risk of negative effects, including psychological effects such as anxiety, paranoia and hallucinations (Hando, Topp and Hall, 1997). These negative effects may also include convulsions, cerebral haemorrhages and respiratory failure (Gay, 1982). Longer duration of use of MA can result in long-term deficits to attention, memory and controlled behaviour and other problems to the cardiovascular system, central nervous system, gastrointestinal system as well as skin and dental problems (Richards and Laurin, 2020). Higher dosages have been linked to enhanced negative effects, addiction, increased neurotoxicity, altered neuroplasticity and higher rates of consumption (García-Cabrerizo and García-Fuster, 2019). Smoking MA was also associated with more frequent usage and similar levels of negative effects when compared to injecting MA (McKetin *et al.*, 2008). The use of MA leads to a rapid release of neurotransmitters, such as dopamine and serotonin, inducing the various positive effects experienced by users as well as tachycardia and elevated blood-pressure (Romanelli and Smith, 2006). Use of MA also results in high neurotoxicity causing damage to the dopamine and serotonin neurons in the brain. Although this damage has been investigated in its relation to the central nervous system, relatively little is known about the underlying molecular mechanisms involved in the neurotoxicity caused by MA abuse (Yu *et al.*, 2015).

1.2 Understanding the molecular underpinnings of Methamphetamine Use Disorder

There is need for improved ways to manage and treat MUD, particularly in South Africa (Lanesman *et al.*, 2019). However, to do this, a greater understanding of the molecular underpinnings of the

disorder is required. While research has implicated several genomic regions and genes associated with MA dependence (Uhl *et* al., 2008) which supports polygenic influences on MUD susceptibility, there is little evidence to explain this when compared to other NP disorders (McClellan, Susser, and King, 2007; Mitchell and Porteous, 2011; Betancur, 2011). Where these studies were able to identify various genes and genomic regions associated with respective disorders such as, COMT, GRM3, G72, *DTNPB1* and *DISC1* in SZ (McClellan, Susser and King, 2007), and *NPHP1*, *RPE65* and *NIPBL* in Autism Spectrum Disorder (ASD) (Betancur, 2011). While these genes are associated with these disorders, they were also found to be associated with several other disorders displaying similar symptoms.

Several studies, including those in twins, have investigated polygenic inheritance in other stimulant class substances, with most of these focusing on CUD (Kendler, Karkowski and Prescott, 1999; Kendler *et al.*, 2005; Hart, de Wit and Palmer, 2012). In comparison, the number of association studies investigating MUD specifically are relatively few when compared to those investigating CUD (Jensen, 2016) or other SUDs (Prom-Wormley *et al.*, 2017). Nevertheless, the studies that do exist have identified several genes associated with inflammation, autophagy and apoptosis (K. Zhang *et al.*, 2016; Kays and Yamamoto, 2019; L. Sun *et al.*, 2019; Wen *et al.*, 2019). A study by Tehrani *et al* (2019) identified dysregulation of apoptotic genes such as *P53*, *BCL2L11*, *BBC3*, *CASP8* and *P21* in rats. The study was also able to identify dysregulation in several genes known to be involved in autophagy and inflammation such as *ATG12 and MAP1LC3B,* and *IL1β*, *IL10* and *HMGB1* related to inflammatory response. These genes are related to multiple molecular processes known to affect the correct functioning of autophagy, inflammatory response, and apoptosis, resulting in neurodegradation (Jayanthi *et al*., 2001; Krasnova *et al*., 2009). A study by Siddiqui *et al*., (2008)focused on investigating expression level changes in the prefrontal cortex which is known to be involved in working memory and the cognitive control of behaviour ; both of which are relevant to MUD aetiology (Zhao *et al.*, 2019).

Several studies have investigated MUD genetic architecture through the use of Genome-wide Association Studies (GWAS), with most of these studies in populations of European or South-East Asian descent (Uhl *et al.*, 2008; Bousman *et al.*, 2009; Ikeda *et al.*, 2013) and as such it cannot be assumed that genomic regions and genes identified in these studies would be achieved with similar significance in local South African populations (Campbell and Tishkoff, 2008). These studies were able to identify a number of genes associated with MA dependence or abuse such as *CDH13*, *CSMD1*, *ABGL1*, *NPAS3* and *DLG2* (Uhl *et al.*, 2008; Ikeda *et al.*, 2013; Jensen, 2016). Several of the genes identified in these studies were also previously associated to other substance use disorders, such as

alcohol use disorder or opioid use disorder (Uhl *et al.*, 2008). While the number of studies focusing exclusively on MUD are limited, there are a few investigating CUD; the findings of these latter studies may be relevant for MUD given that MA and cocaine are both stimulants (American Psychiatric Association, 2013).

To our knowledge there is no literature on the molecular underpinnings of MUD in a South African context. This lack of literature makes identifying biomarkers that could lead to better treatments or therapies challenging. This is further exacerbated by the fact that there is little research that is able to depict the difference, molecularly and genetically, between MUD and other SUDs despite MUDs classification as a stimulant use disorder in the DSM-5 (American Psychiatric Association, 2013). This leads to further problems considering that many SUDs share pathophysiology but present clinically distinct phenotypes, making it hard to devise more precise treatments for these individuals. Newer study approaches utilizing GWAS and elaborating on the data through *in silico* or bioinformatic analyses may aid in fine-mapping and characterising the underlying molecular underpinnings of the disorder.

1.3 Premise and pitfalls of genome wide association studies

The employment of GWAS approaches is aimed at unveiling causal mechanisms underlying the relationship between common genetic variation and disease, biological characteristics, or drug response (Need and Goldstein, 2010). The premise of this approach lies in a lack of ambiguity in performing the study devoid of prior knowledge of the genetic basis of the specified disorder/s under investigation (Simón-Sánchez and Singleton, 2008; Need and Goldstein, 2010). This allows for investigation of individually targeted disorders, not being resultant of a singular causal gene variant of strong influence (Need and Goldstein, 2010). GWAS protocols utilize arrays of common single-nucleotide polymorphisms (SNPs), taken as representative of the entire genome (Need and Goldstein, 2010). In theory this relates to potential for the entire genome to be studied as "targeted" SNPs infer information regarding various other non-genotyped variants (Need and Goldstein, 2010). This inference is on the basis of linkage-disequilibrium (Need and Goldstein, 2010). As is apparent from large GWAS results, the identification of causal genes not previously implicated or suspected in disease aetiology, can now be identified (Stranger, Stahl and Raj, 2011). GWAS is thus not as restricted as candidate gene studies in which a candidate gene/s are specifically selected for further inclusion and investigation. GWAS may also be able to estimate genetic effects (additive and non) and pleiotropy (Stranger, Stahl and Raj, 2011). Furthermore, subsequent studies have identified that for almost any complex trait, variants tend to have small effect sizes. GWAS data has been indicative that most disorders are consequential of hundreds of common variants implicated with associated

SNPs; however, these results explain only a small portion of genetic risk (Cantor, Lange and Sinsheimer, 2010), and have only lead to correlation and not causation in this regard. In addition, GWAS still ignore the fundamental principles of gene expression, both in mechanism of action and genetic influence that can alter ordinary gene expression (Lahiri and Maloney, 2012). This comes about due to GWAS presuming that an individual's state of health is determined by a fixed genome, when this is not the case as a genome's interaction with the environment changes the expression of this genome and subsequently the individual's state of health (Lahiri and Maloney, 2012). Despite a large number of genomic associations having been found utilising various GWAS study designs leading to the elucidation of underlying genetic factors of respective neuropsychiatric disorders, this tool has come short in representation of genetic diversity, appropriate statistical power to find associations and incorporation of external (environmental) factors influencing genomic variation (Popejoy and Fullerton, 2016). This deems the GWAS approach problematic when taking into consideration the genetic diversity and genomic admixture pertaining to individual population groups and genomic and phenotypic differences observed (Ramsay *et al.*, 2011; Quansah and Karikari, 2015). In this manner bias is introduced inadvertently towards the population group under study. Study designs inclusive of particular population groups (i.e. African) will find GWAS results that differ vastly from that of more 'commonly' employed population groups making up cohorts (e.g. of European decent), due to differences in linkage blocks (Quansah and McGregor, 2018).

Although the GWAS protocol itself initially seems to be a non-biased approach, the manner in which SNPs are taken as 'representative' of the entire genome is naïve in that these SNPs cannot be relied on for validity of samples outside of the study at hand. Different population groups are likely to have different linkage blocks and minor-allele frequencies (MAF), confounding association results (Tan, 2007). Furthermore, associated significant SNPs identified by GWAS are largely found within non-coding regions of the genome (The Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium, 2011). The discovery of disease associated variants in intronic regions of the genome has led to an increase in studies looking at non-coding variants with a lot of traction being gained through the study of miRNA in NP disorders due to their environmental interaction (O'Connor *et al.*, 2016). Thus, while the use of GWAS on its own is unable to provide all the information it initially promised, the use of other tools such as imputation, enrichment, and various statistical analyses can further enhance our understanding of the data.

1.4 miRNA and Their Potential as Biomarkers in Complex Disorders

miRNA are small, 22 nucleotide (nt) RNA structures that are able to regulate gene expression through binding to the 3' untranslated region (UTR) of messenger RNA (mRNA) (Krol, Loedige and

Filipowicz, 2010). There are constantly other regulatory mechanisms being discovered, with recent studies estimating that 59% of binding sites lie outside of the 3'UTR (Boudreau *et al.*, 2014). Interactions between genetic architecture and environmental factors have long been suspected to be associated with psychiatric morbidity, and in triggering the onset of such disorders (Issler and Chen, 2015). Environmental factors can induce changes in gene expression levels, which in turn mediates the onset of disorders, albeit with no alteration of the DNA sequence (Jirtle and Skinner, 2007; Cortijo *et al.*, 2014). miRNA were first discovered by Lee et. al. (1993), when it was found that the gene *LIN4*, in *Caenorhabditis elegans*, encoded an RNA transcript complementary to the sequence of *LIN14*. Further experimentation led to the discovery that elevated *LIN4* levels had a negative effect on *LIN14* levels (Wightman, Ha and Ruvkun, 1993). Since then, miRNAs have been identified as key eukaryotic gene expression regulators with an estimated 60% of mammalian mRNA transcripts being miRNA targets (Friedman *et al.*, 2009). Mature miRNAs are derived from the transcription of a primary miRNA transcript (pri-miRNA) which can be over 1kb in length and are usually characterised by a stem-loop structure (Luoni and Riva, 2016). These transcripts are then processed into precursor miRNA (pre-miRNA) which are between 60 and 100nt long (Lee *et al.*, 2003). Pre-miRNAs are translocated out of the nucleus via Exportin-5 interaction (Yi *et al.*, 2003). Once released from the receptor, it is further processed to generate an unstable ~22nt duplex which produces two single-stranded RNA species, one which is preferentially degraded and one which is incorporated into the RNA-induced silencing complex (RISC); the miRNA processing is performed by two RNases, i.e. Drosha and Dicer (Luoni and Riva, 2016). The entirety of this process is called miRNA biogenesis and can be visualized in Figure 1. miRNA binding does not require perfect binding; a single miRNA is able to bind many different mRNAs (Martinez and Walhout, 2009). Inversely, mRNA may also have more than one miRNA binding site and as such may be bound by more than one miRNA resulting in numerous combinations.

The regulation of the miRNA target is dependent on the level of complementarity between the miRNA and the mRNA target and will either lead to repression or degradation. miRNA expression is regulated much in the same way as mRNA, as most of the transcription factors (TFs) that regulate mRNA expression also regulate miRNA expression (Luoni and Riva, 2016). However, miRNA transcription is interesting in that it acts as a feedback loop whereby a miRNAs downstream target could be a TF acting on its transcription (Martinez and Walhout, 2009). An example of this mechanism of miRNA regulation is the transcriptional repressor REST which is able to directly regulate the transcription of brain-related miRNAs, for example, miRNA-124a and miRNA-132,

which control the perseverance of neuronal transcripts, determining cellular phenotype (Luoni and Riva, 2016).

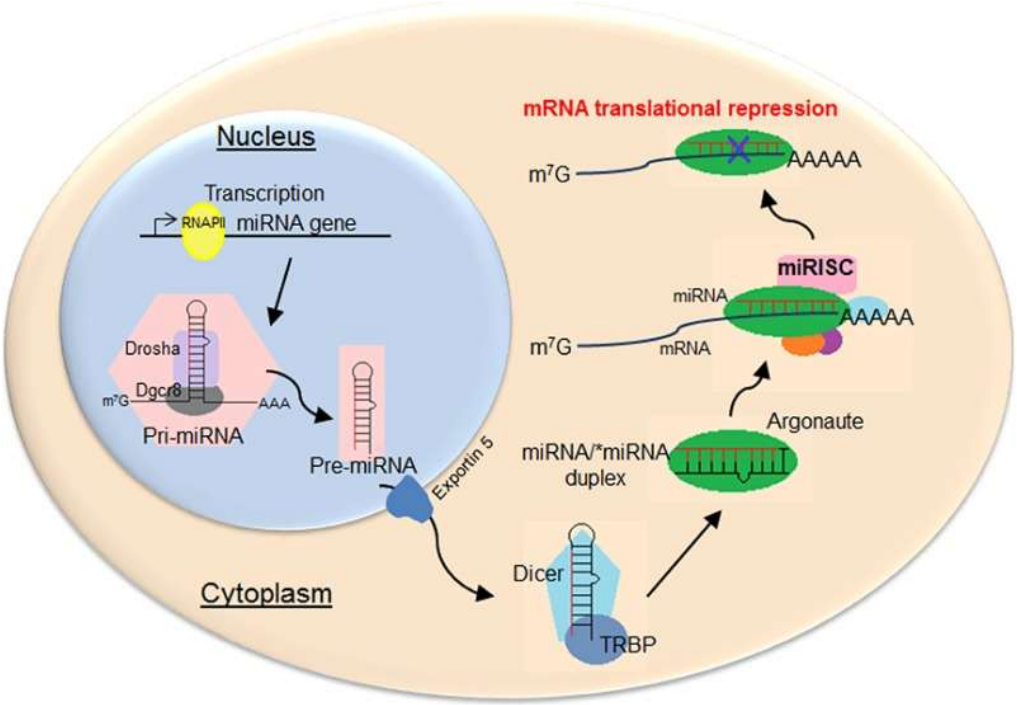


Figure 1: Figure from (Hajarnis, Lakhia and Patel, 2020) indicating the process of miRNA biogenesis from pri-miRNA to the mature miRNA complex. This indicates the start of biogenesis in the nucleus with the transcription of pri-miRNA sequence before processing into the stem-looped pre-miRNA structure that is transported out of the nucleus. Finally, the pre-miRNA is further processed into mature miRNA complex via Dicer before associating with the RISC complex to form the complex that regulates mRNA translation.

Numerous studies have suggested a role for miRNAs in almost all biological processes ranging from cell proliferation, differentiation, development, and apoptosis (Krol et al., 2010). More significantly, miRNAs have been shown to have important roles in central nervous system development and function with regards to neuronal development and synaptic plasticity (O'Connor et al., 2016). It has also been shown that the deregulation of miRNA expression levels and function is associated with the pathogenesis of NP disorders (Alural et al., 2016). Studies identifying miRNA effects used this knowledge as disruption of the miRNA processing pathway (using Dicer and Drosha) affects global miRNA functionality and thus allows for the observation of miRNA effects on an organism at a global scale. This is exemplified by studies showing that the removal of Dicer led to decreases in the development and maintenance of the nervous system in cell culture (Barbato *et al.*, 2007; Schaefer *et al.*, 2007). Synaptic plasticity has also been shown to be influenced by miRNA regulation as it

requires a functional RISC pathway for synaptic protein synthesis (Ashraf *et al.*, 2006). Many of the miRNA targets in the brain are considered risk genes for the development of disorders like autism spectrum disorder (ASD), schizophrenia (SZ), bipolar disorder (BD) and major depressive disorder (MDD) (Camkurt et al., 2017). Thus, it can be clearly discerned that miRNA regulation is an important factor in correct neurodevelopment and maintenance of healthy neural functioning.

Studies in rats have shown that some therapeutics are able to alter miRNA expression. A study on mood stabilizers has shown the differential expression of several miRNAs in the hippocampus predicted to regulate gene products associated with BD. This study provided the first evidence indicating miRNA involvement in psychiatric disorders and in successful therapeutics (Zhou *et al.*, 2009). Since the study, other pharmacotherapeutics have been shown to have miRNAs as downstream targets. Fluoxetine, for example, is a common therapeutic used to treat depression and works by inhibiting the serotonin transporter (SERT) (Baudry *et al.*, 2010). Baudry et al. (2010) indicated that the inhibition of SERT by fluoxetine is through increasing the levels of miRNA-16 which targets SERT in serotonin raphe neurons as it was also observed that directly increasing miRNA-16 in raphe neurons displayed similar results to that of fluoxetine treatment. Studies have also shown that miRNA biogenesis in the nucleus accumbens is highly important for the generation of behavioural resilience to social-defeat stress, as Dicer1-deficient animals were particularly affected by milder defeat procedures when compared to wild-type (WT) animals, which were unaffected (Dias *et al.*, 2014). There is increasing support for the claim that changes to miRNA function may act as the link between psychological stress and downstream pathophysiology (O'Connor *et al.*, 2016). Early life adversity (ELA) in rodents was used to induce a state of hypersensitivity to later, chronic, unpredictable mild stressors which, in turn, correlated to a decrease in striatal miRNA-9 (Zhang *et al.*, 2015). This decrease reflected the downstream repression of the dopamine repressor D2. Thus, from their role as a mediator in stress-induced pathology, it may be possible to target miRNAs in order to affect their downstream targets as a form of therapy (Zhang *et al.*, 2015).

miRNA involvement in the molecular pathophysiology of psychiatric disorders makes them a viable candidate as a diagnostic biomarker if significant differences can be identified between cases and healthy controls (O'Connor et al., 2016). Currently, there is strong evidence to support miRNA use in diagnosis as there have been distinct differences in miRNA expression levels between patients with depression and controls (O'Connor et al., 2016). It was also found (Garbett et al. 2015) that there is a high correlation between levels of specific miRNAs and their target mRNAs in these same cellular samples. While it may be true that miRNA involvement in neuronal functioning does not always predict a coinciding difference in the same mRNAs in blood, there is mounting evidence that this is

the case for some miRNAs (Lopez *et al.*, 2014, p. 1202). There is potential that miRNAs may serve as an adequate form of diagnosis for several neuropsychiatric disorders, including SUDs. In future, the use of miRNA as biomarkers could lead to more specialized medicine and molecular-based treatments based on detected biomarkers indicative of specific diagnostic criteria or treatment response (Belzeaux, Lin and Turecki, 2017; Swarbrick *et al.*, 2019). Furthermore, it is possible that with further research in diverse populations miRNA could be used as a non-invasive screening tool to potentially diagnose individuals at early stages of disorder progression, with possible miRNA biomarkers already being identified for disorders such as SZ and MDD (Alural, Genc and Haggarty, 2017)

1.5 The role of miRNA in Methamphetamine Use Disorder

The missing information on heritability of SUDs has led to a shift in focus towards the underlying mechanisms to which most GWAS results speak, in the hopes of uncovering common mechanisms. There have been several studies examining the influence of environmental effects, on miRNA and the associated biological pathways (Dickson *et al.*, 2018; Cattane *et al.*, 2019). These studies indicated that environmental effects like childhood trauma can affect miRNA expression levels and that these changes in expression can, in some cases, lead to lasting effects over generations. This is exemplified in the study by Dickson *et al* (2018) in which miR-499 and 34 were found to have lower expression levels in individuals with a higher rating of early-life adversity, with the changes in these same miRNA in mice found to carry across generations. A study by Catrane *et al* (2019) found downregulation of miR-125b-1-3p was associated with early life stress and enhanced susceptibility to schizophrenia development. These studies indicate the potential for environmental stressors to affect miRNA and subsequent regulatory targets of miRNA known to be associated with neuropsychiatric disorder aetiology.

Despite the above-mentioned complex nature of MUD, studies investigating it have identified several of candidate genes from clinically diagnosed patients with genome-wide studies utilizing computational approaches. These study approaches have conceptualised hypotheses of the potential underpinning molecular mechanisms and involved pathways in MUD and other NP disorders (Cristino *et al.*, 2014). Functional studies, investigating the biochemical, cellular and physiological roles of miRNA, have identified roles of miRNAs in controlling cellular processes, like neurogenesis and synaptic plasticity, with both mechanisms having been implicated in NP disorders (Kloosterman and Plasterk, 2006; Magill *et al.*, 2010; Saba and Schratt, 2010; Bredy *et al.*, 2011; Luikart, Perederiy and Westbrook, 2012). Furthermore, there have been studies that identified miRNAs associated with multiple disorders; this indicates that while there may be unique clinical manifestations, the overlap

of miRNA and subsequent genes and biological pathways speaks to the overlapping nature of these disorders and frequent comorbidity (Mendes-Silva *et al.*, 2016). Thus, it can be clearly discerned that miRNA-mediated regulation is an important factor for neurodevelopment and maintenance of neuronal functioning, within the context of these overlapping NP disorders.

Despite this, there have been relatively few studies investigating the involvement of miRNA in the molecular underpinnings of MUD specifically. Of these studies, a large number were done using animal studies or bioinformatic prediction tools (Zhang *et al.*, 2016; Sim *et al.*, 2017; Li *et al.*, 2018). For example, the study by Li *et al* (2018) identified several differentially expressed miRNA in rats, in response to MA use which were associated with biological processes related to synaptic transport mechanisms, apoptotic cell death and neurogenesis; all of which are processes known to be involved in MUD disorder aetiology (Zhang *et al.*, 2016; Li *et al.*, 2018; Sun *et al.*, 2019). The study by Sim *et al* (2017) also investigated differential expression of miRNA in rats, identifying several miRNA, notably miR-496-3p, miR-194-5p, miR-200b-3p and miR-181-5p, which displayed significant associations with MA addiction. Lastly, the study by Zhang *et al* (2016) was able to identify a miRNA-mediated interaction with *BBC3* expression related to apoptosis and autophagy regulation; both processes with major involvement in MA neurotoxicity (Zhang *et al.*, 2016; Sun *et al.*, 2019).

Considering the link between environment, miRNA, and NP disorders, it seems highly plausible that pathways implicated in NP disorders could be regulated differently resulting in differing clinical manifestations. This begs the question whether it is possible that there is a common sub-component that is being regulated differently in multiple disorders, causing them to present distinct clinical manifestations. And is there a difference, in terms of molecular underpinnings, between MUD and other subtypes of SUD? To answer these questions, more studies focusing on the role of miRNA-mediated and epigenetic regulation in MUD are needed.

In conclusion, due to the burden associated with MUD and the limited amount of clinically relevant information obtained from GWAS alone, studies have started to investigate other mechanisms that may be involved in the molecular underpinnings of the disorder, such as miRNA. Environmental stressors can affect miRNA-related regulation of the genome and exert effects on disorder aetiology. This evidence along with the lack of clinically relevant information on MUD using GWAS has led to studies investigating the role of miRNA in MUD. Most of these studies were either performed in homogeneous human populations or in animal studies, however, investigating this in a diverse South African cohort with MUD to increase statistical power and using enrichment to investigate biological pathways along with better characterised discovery groups could result in novel data on associations as well as revalidate previous findings.

Thus, it is the aim of this study to further elucidate the potential regulatory role of miRNA in MUD by using a PGC-CUD discovery cohort alongside a local MUD cohort. This will be accomplished by identifying CUD genes and miRNA via an *in silico* bioinformatic pipeline. The frequency of these loci will be characterised in the local MUD cohort. Prioritised MUD variants will be identified by performing association analyses before being enriched for any miRNA-mediated regulatory potential using *in silico* online bioinformatic tools. The data of both datasets will then be compared to identify common and unique loci in the MUD cohort. This would indicate the molecular differences between MUD and other SUDs while providing insight into the regulatory network of MUD.

Chapter 2 Materials and Methods

This study aimed to identify the potential regulatory role of miRNA in MUD using an exploratory methodology (i.e. no *a priori* hypothesis). The objectives were: 1) to do a case-control GWAS of a local MUD cohort using logistic regression analysis, 2) to conduct a bioinformatic analysis of a CUD discovery cohort and the MUD GWAS cohort, 3) to do an enrichment analyses of host genes identified in each cohort and 4) to do a comparison of the bioinformatic profiles of each cohort.

Role of the incumbent

Clinical interviews, patient treatment, data collection and blood/saliva sampling were performed by trained clinicians. DNA extractions were performed by a laboratory technician and the incumbent. Genotyping was performed by the Broad Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America. The role of the incumbent of this study was therefore to make use of existing genotype data of a CUD discovery cohort (Gelernter *et al*., 2014). In addition to this, the incumbent was to generate new GWAS data for a local MUD cohort by preparing DNA samples and sending to the Broad Institute for GWAS analyses using the Illumina Global Screening Array (GSA). The incumbent wrote and executed scripts to conduct quality control and binomial logistic regression analyses to identify SNPs associated with MUD. Data was subsequently mined from each cohort for use in an *in silico* bioinformatic pipeline (which the incumbent generated) to identify genes and miRNA of interest. The incumbent then collated this data and enriched it using online tools in line with the aim and objectives of the study. The results were discussed, and conclusions drawn by the incumbent, under supervision.

2.1 Cohort demographics

This study made use of two cohorts: a CUD discovery cohort consisting of summary statistics data from 5697 individuals, and a South African cohort consisting of data from 47 adult patients with MUD and 34 age and sex-matched controls. Blood or saliva samples were obtained from South African MUD patients recruited for another project (Lanesman *et al* 2019). Patients were recruited from in- and outpatient rehabilitation centres and the community. DNA was subsequently extracted as described by Miller, Dykes and Polesky (1988) for blood samples and as per Oragene-DNA OG500 tubes (DNA Genotek, Ontario, Canada) for saliva samples. Two cohorts were used to identify genes that were both common between disorders and/or unique to the MUD cohort. The individuals with MUD were from mixed-ancestry and their age and sex are described below in Table 2, whereas the CUD discovery cohort has been previously described by Gelernter *et al.*, (2014). The mixed-ancestry cohort was used in order to increase the power to detect association and to increase the potential to identify novel associations; this is due to the fact that the mixed-ancestry cohort is heterogeneous

when compared to MUD cohorts previously studies (European or Asian descent). While some studies have identified higher rates of MA use in South African mixed-ancestry in community-based studies (Watt *et al.*, 2015), other studies have found that this is not the case (Weybright *et al.*, 2016) indicating that more work investigating the incidence demographics needs to be conducted.

Table 2: Cohort demographics for the South African methamphetamine use disorder GWAS.

Characteristic	All (n = 81)	Case (n=47)	Control (n=34)
Demographic (Age ± SD)			
Male sex	31 (39 ± 9)	18 (38 ± 11)	13 (37 ± 11)
Female sex	50 (34 ± 9)	29 (37 ± 10)	21 (37 ± 12)

2.2 Genome-wide association analysis of MUD cohort

Genome-wide association screening of DNA, extracted from blood and saliva for the MUD cohort, was performed using the Illumina GSA v24 (https://www.illumina.com/science/consortia/human-consortia/global-screening-consortium.html). Data files were processed at the Broad Institute, Stanley Center (https://www.broadinstitute.org/stanley).

2.3 Imputation of MUD GWAS data

Imputation was performed using the Sanger Imputation Service (https://imputation.sanger.ac.uk/) which allows for the use of SHAPEIT2 (http://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html) phasing and the African Genome Resources haplotype reference panel (https://www.apcdr.org/data/) which has been shown to be useful for the local population used in the MUD cohort (Schurz *et al.*, 2019). Several checks were made to ensure the data followed the imputation server's guidelines before the data was uploaded and imputed. Imputed data was filtered using an info score threshold of 0.3 (Marchini and Howie, 2010; Roshyara *et al.*, 2014; Eller, Janga and Walsh, 2019). Imputed data was then updated with individuals' sex and phenotypic status (case or control) before quality control was performed.

2.4 Quality control of MUD GWAS data

Quality control was performed using Plink v1.9 (Chang *et al.*, 2015) as performed by Anderson and Clarke (2010), with minor adjustments made for the sample size of the study and to match the quality control performed on the discovery cohort. Gelernter *et al* (2014) excluded individuals with call rates (MIND and GENO) <98% with a minor allele frequency (MAF) threshold of 1% whereas our data was filtered according to minor allele frequency (MAF > 0.05) and call rates per marker and

individual <95% (genotyping rates) (MIND and GENO 0.05). The data was also filtered for Hardy-Weinberg equilibrium (HWE; $p > 1x10^{-6}$) and sex mismatches. These quality control steps were performed in order to identify individuals with discordant sex classifications, missing genotype or heterozygosity rates, duplicated or related individuals and ethnic outliers; these quality control steps also identified SNPs deviating from HWE, low MAF and significantly different missing genotype rates. All quality control checks were performed using Plink and were visualised using R (R Core Team, 2019) to produce histogram and Q-Q plots where relevant via the *qqman* package (Turner, 2014). These plots check the distribution of MAF while also comparing the expected outcomes of HWE and individual and genotyping call rates to the observed outcomes. Principle component analysis as described by Anderson and Clarke (2010) was performed using SMARTPCA (Price *et al.*, 2006), however, this further reduced cohort size and limited the statistical power to detect association; thus ethnic outliers were included in the study and results were interpreted with this in mind. Thresholds were selected and adjusted according to study size and thresholds proposed by Gelernter *et al* (2014).

2.5 Association analysis of MUD GWAS data

A case-control binomial logistic regression analysis was performed using Plink on GWAS data passing quality control filters using phenotype and genotype as the binary variables for association testing. A binomial logistic regression assumes that the observation will fall into a binary category based on one or more independent variable that is either continuous or categorical; in this study the dependent variable, phenotype, is binary (case or control) and the categorical variable is genotype, with no covariates in order to maintain maximum statistical power. The association was visualised using R to produce Manhattan and Q-Q plots to visualise SNP locations and p-values, as well as too assess the reliability of the data by comparing expected p-values to the observed p-values. A power calculation was performed using CaTS (Center for Statistical Genetics, University of Michigan, http://www.sph.umich.edu/csg/abecasis/CaTS/tour1.html). To achieve 80% power, assuming a significance of 0.05, cohorts of much larger size than that used in this study would be needed. Considering the cohort size in this study and a prevalence rate of 0.3 for MA dependence in South Africa (Peltzer and Phaswana-Mafuya, 2018), a 75% power can be achieved assuming a false-positive maximum of 3 for the SNPs investigated for a relative genotype risk of 1.5 (referring to common polymorphisms with a MAF ≥ 0.2) (Skol *et al.*, 2007).

2.6 CUD summary statistics *in silico* analyses

The summary statistics for the most recent CUD dataset were obtained from the PGC-SUD workgroup. Single nucleotide polymorphisms (SNP(s)) passing a nominal significance threshold of

$p \leq 1x10^{-4}$ were prioritised from the data. This nominal significance threshold was used in lieu of the traditional genome-wide significance threshold of $p \leq 5x10^{-8}$ as even studies performed by large consortia struggle to identify SNPs passing such a stringent threshold (Gelernter *et al.*, 2014). As such, this study chose to use a threshold commonly utilised in smaller studies to identify SNPs approaching significance that may reach significance when investigated in a large enough cohort (X. Wang *et al.*, 2016; Stahl *et al.*, 2018; McGregor *et al.*, 2019). Once prioritised, the SNPs were input into an *in silico* bioinformatics pipeline using online tools, namely SNPnexus (http://snp-nexus.org/; accessed August 2019), TarBase (http://carolina.imis.athena-innovation.gr/diana_tools/web/; accessed August 2019) and Genetic Association Database (https://geneticassociationdb.nih.gov/; accessed August 2019) in order to investigate miRNA-related gene regulation and related host genes (Becker *et al.*, 2004).

SNPnexus assesses known and novel variants' significance in relation to the transcriptome, proteome, regulome and structural variation through the use of multiple third-party databases pointing to altered gene, protein or regulatory isoforms that could lead to phenotypic changes of interest. This makes aggregating variant annotation information from these various databases in a single query possible instead of requiring researchers to investigate the variants using each database individually (Ullah *et al*., 2018).

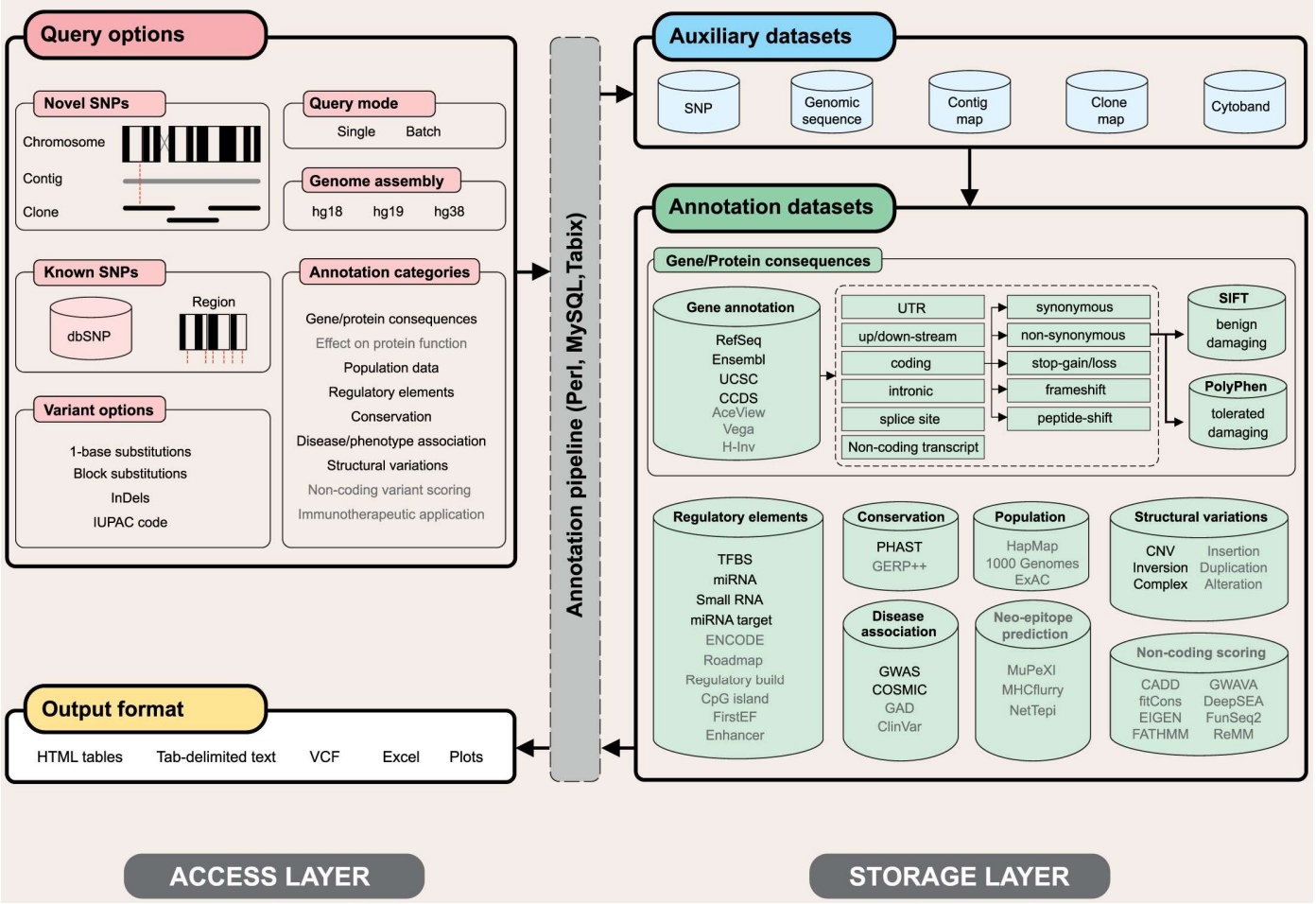


Figure 2: Depiction of the process through which SNPnexus operates. An initial query is made using chromosomal information or known SNP identities before being processed by the auxiliary datasets. Subsequently, information is collated from annotation datasets selected by the user and compiled into the chosen output format (Dayem Ullah *et al.*, 2018).

TarBase is an online database containing information on over 670 000 experimentally validated miRNA-gene interactions across various cell types and tissues. Later versions of the TarBase tool are integrated with ENSEMBLv99 (https://www.ensembl.org/index.html) (Hunt *et al.*, 2018) allowing for the viewing of the exact binding location using ENSEMBL's genome browser, as well as containing an integrated ranking system that displays interactions based on the robustness of the methodologies that validated them. The ranking system determines the interaction's ranking based on the experimental validation method used (either low- or high-throughput experiments) (Karagkouni *et al.*, 2018).

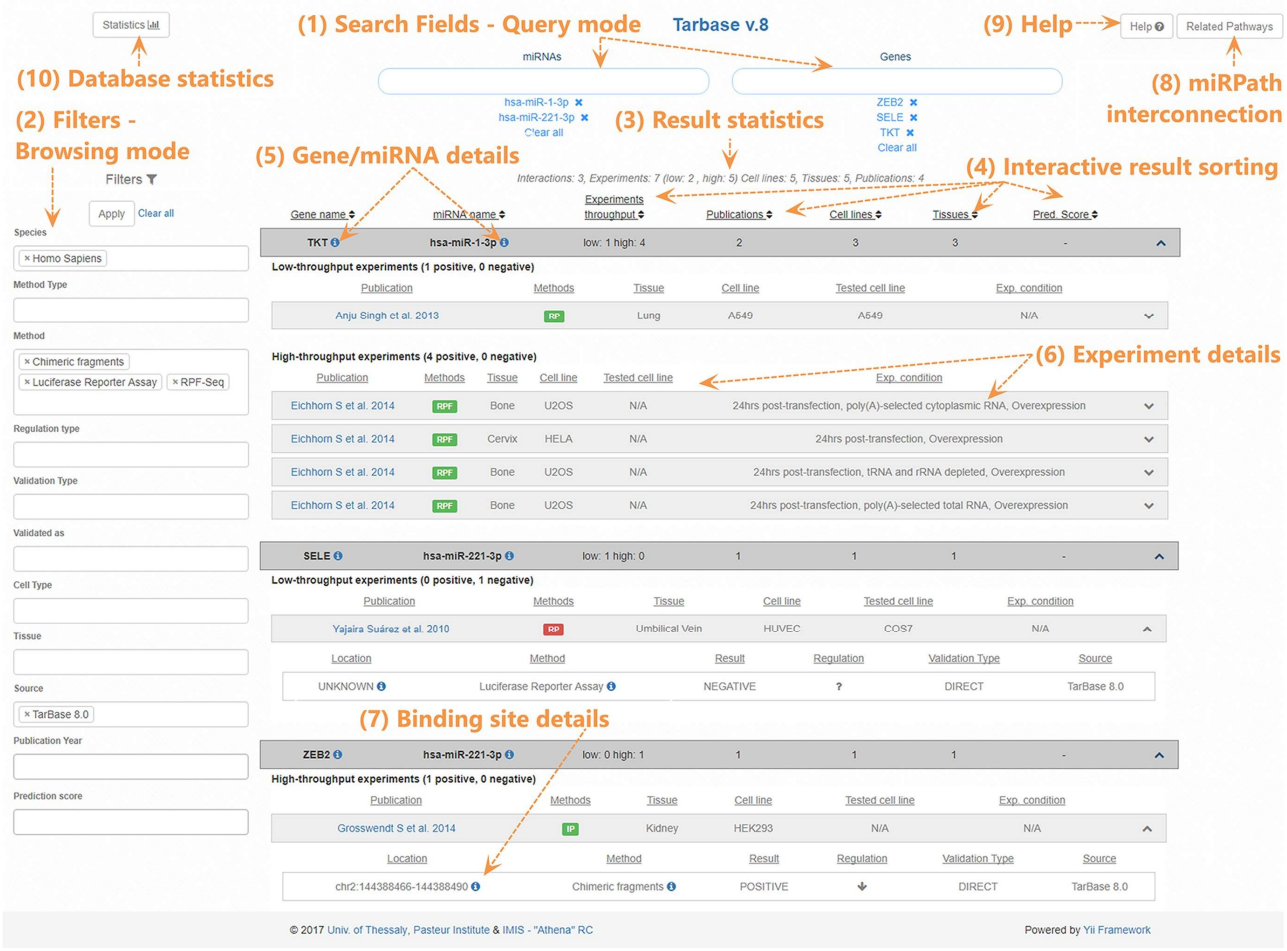


Figure 3: Image depicting the DIANA-TarBase v8.0 interface. Queries can be made using either miRNA and/or gene names or by navigating through database content using criterion filters. Interactions can be further refined using filtering options for species, tissues/cell types, methodologies, type of validation, source, etc. Gene and miRNA details are complemented with active links to Ensembl, miRBase and the DIANA disease tag cloud. Interactions are also accompanied by miRNA-binding site details (Karagkouni *et al.*, 2018).

Genetic Association Database is a database containing genetic association data pertaining to complex diseases and disorders. This database was designed to collect, standardize, and archive genetic association data and allow for easy access via online tools or direct download (as of 2014 when all data was "frozen"). The database standardizes the nomenclature of the data input as well as annotates each record with links to relevant molecular databases and reference databases. The data is annotated to include the official gene symbol (HUGO gene symbols), disease phenotype and class, gene-based molecular information, chromosomal and mutation information, and relevant references (Becker *et al.*, 2004).

Subsequently, the host genes associated with the prioritised SNPs identified using GAD were used for gene enrichment analyses. Host genes for the identified SNPs were input into the *in silico* online tool, Enrichr (https://amp.pharm.mssm.edu/Enrichr/; accessed August 2019) (Chen *et al.*, 2013). Enrichr uses 35 gene-set libraries investigating six categories (transcription, pathways, ontologies, diseases/ drugs, cell types and miscellaneous) to compute enrichment. Enrichment is calculated using Fisher's exact test assuming a binomial distribution and independence for the probability of any gene belonging to any set. Corrected p-values are calculated by computing a Z-score statistic where the mean rank and standard deviation from each computed rank calculated for each term in a gene-set library where the Z-score represents the deviation from this expected rank (Chen et al., 2013). The most accurate statistical representation is defined by the combined score which is a representation of the Fischer exact and Z-score statistics (Chen et al., 2013). Correction for multiple testing was accounted for using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995).

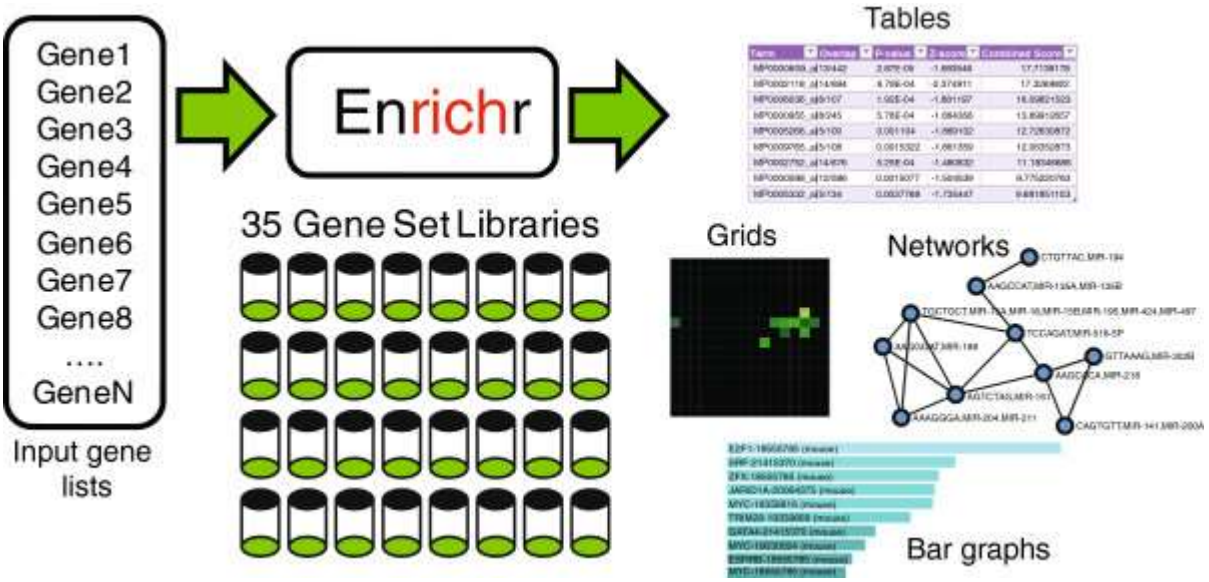


Figure 4: Figure depicting the workflow by Enrichr. Input genes are analysed as mentioned above using the 35 gene-set libraries with enriched data being collated in a series of tables, grids, networks, and bar graphs (Chen *et al.*, 2013).

2.7 MUD GWAS *in silico* analyses

After the association analysis had been carried out, SNPs were filtered for a nominal significance value of $p<1x10^{-4}$. SNPs passing this threshold were used in the in *silico* bioinformatic pipeline (SNPnexus, TarBase and Enrichr).

2.8 Cohort comparison

Following the *in silico* analyses of both cohorts, the gene lists between the two cohorts were compared for common and unique genes in the MUD cohort by identifying genes associated only with the MUD cohort instead of genes identified in both cohorts. Once any host genes exclusive to the MUD cohort were identified, these underwent enrichment analyses again to identify biological pathways and miRNA-regulatory interactions associated with these genes to better characterise the molecular difference between the two cohorts.

Chapter 3 Research Results

3.1 Demographics

The MUD cohort consisted of cases and controls that were age and sex matched (Table 2). These individuals' data underwent standard GWAS quality control as per Anderson and Clarke (2010); no individuals were removed from the study. As this was a case-control study design, logistic regression parametric analyses were used for this cohort. The CUD discovery cohort underwent analyses as described by Gelernter *et al* (2014) and only summary statistics were used for the subsequent *in silico* analyses.

3.2 Discovery (CUD) cohort description

The CUD discovery cohort was comprised of summary statistics from a study by Gelernter *et* al (2014). Prioritisation of the CUD summary statistics yielded 807 SNPs which presented with a nominal significance of $1x10^{-4}$ or lower (Supplementary Table 1). These SNPs were used in the *in silico* bioinformatics pipeline described above.

3.3 Discovery (CUD) cohort *in silico* analyses

The SNPs were input into SNPnexus and TarBase to identify any SNPs involved in direct miRNA-mediated regulation via affecting target sites or miRNA encoding sites. Of the 807 SNPs prioritised, three presented with direct miRNA-mediated regulation potential when investigated using TarBase and SNPnexus (Table 3).

Table 3: SNPs involved in miRNA-mediated regulation based off investigation via SNPnexus and TarBase using the CUD summary statistics. TarBase outputs the data relating to affected target sites as well as strand information (where 1 is forward strand and -1 is reverse strand) for the investigated SNPs and associated miRNA.

SNP	Target Site	Strand	miRNA
rs2073900	19:35250535-35250583	1	hsa-let-7e-5p
rs4598	6:36494228-36494249	-1	hsa-miR-196b-5p
rs4598	6:36494228-36494250	-1	hsa-miR-196a-5p
rs4598	6:36494229-36494250	-1	hsa-miR-140-5p
rs2239808	6:36479194-36479218	1	hsa-miR-210-3p

Secondly, the SNPs were investigated using SNPnexus and GAD which yielded GAD disease class, phenotype, and host genes (n = 150) (Supplementary Table 2). The host genes identified via GAD were used in subsequent enrichment analyses.

Subsequently, enrichment analyses were performed on the identified host genes using the Enrichr online bioinformatic tool, yielding the top 50 gene ontology (GO) results for the 150 host genes (Table 4). Investigation of miRNA-mediated regulation potential via Enrichr was performed for the 150 host genes, which were predicted to be regulated by over 50 miRNA transcripts independent of summary statistics-prioritised SNPs (Table 5). While not all predictions made by Enrichr may be biologically relevant to the study, Enrichr ranks the outputs by statistical significance (Z-score), as described previously (Chen *et al.*, 2013).

Table 4: Top 50 GO biological processes associated with CUD host genes identified via Enrichr. Data output by Enrichr is listed according to Z-score as this is the most accurate measure of association, as described above (Chapter 2, page 20).

Term	P-value	Adjusted P-value	Odds Ratio	Z-Score
vocal learning (GO:0042297)	8,22E-04	1,00	44,44	315,74
imitative learning (GO:0098596)	8,22E-04	1,00	44,44	315,74
calcium ion transport into cytosol (GO:0060402)	3,67E-04	0,94	21,05	166,54
positive regulation of cation channel activity (GO:2001259)	8,17E-05	0,42	17,20	161,94
neuron cell-cell adhesion (GO:0007158)	4,79E-03	1,00	19,05	101,73
myeloid dendritic cell activation (GO:0001773)	4,79E-03	1,00	19,05	101,73
positive regulation of calcium ion transmembrane transporter activity (GO:1901021)	7,06E-03	1,00	15,69	77,70
regulation of B cell receptor signalling pathway (GO:0050855)	7,06E-03	1,00	15,69	77,70
positive regulation of action potential (GO:0045760)	4,42E-02	1,00	22,22	69,33
regulation of protein localization to early endosome (GO:1902965)	4,42E-02	1,00	22,22	69,33
hindlimb morphogenesis (GO:0035137)	4,42E-02	1,00	22,22	69,33

polyamine biosynthetic process (GO:0006596)	4,42E-02	1,00	22,22	69,33
regulation of endothelial cell development (GO:1901550)	4,42E-02	1,00	22,22	69,33
negative regulation of chromosome organization (GO:2001251)	4,42E-02	1,00	22,22	69,33
negative regulation of transforming growth factor beta production (GO:0071635)	4,42E-02	1,00	22,22	69,33
regulation of transforming growth factor beta2 production (GO:0032909)	4,42E-02	1,00	22,22	69,33
positive regulation of protein localization to endosome (GO:1905668)	4,42E-02	1,00	22,22	69,33
synaptic transmission, glycinergic (GO:0060012)	4,42E-02	1,00	22,22	69,33
folic acid-containing compound biosynthetic process (GO:0009396)	4,42E-02	1,00	22,22	69,33
positive regulation of aspartic-type peptidase activity (GO:1905247)	4,42E-02	1,00	22,22	69,33
positive regulation of protein localization to early endosome (GO:1902966)	4,42E-02	1,00	22,22	69,33
mRNA 3'-splice site recognition (GO:0000389)	4,42E-02	1,00	22,22	69,33
regulation of T cell tolerance induction (GO:0002664)	4,42E-02	1,00	22,22	69,33
indolalkylamine metabolic process (GO:0006586)	4,42E-02	1,00	22,22	69,33
positive regulation of podosome assembly (GO:0071803)	4,42E-02	1,00	22,22	69,33
ceramide transport (GO:0035627)	4,42E-02	1,00	22,22	69,33
positive regulation of sodium ion transmembrane transport (GO:1902307)	4,42E-02	1,00	22,22	69,33
regulation of anion channel activity (GO:0010359)	4,42E-02	1,00	22,22	69,33
cellular response to magnesium ion (GO:0071286)	4,42E-02	1,00	22,22	69,33
negative regulation of delayed rectifier potassium channel activity (GO:1902260)	4,42E-02	1,00	22,22	69,33

regulation of epithelial to mesenchymal transition involved in endocardial cushion formation (GO:1905005)	4,42E-02	1,00	22,22	69,33
spermine metabolic process (GO:0008215)	4,42E-02	1,00	22,22	69,33
insulin secretion involved in cellular response to glucose stimulus (GO:0035773)	4,42E-02	1,00	22,22	69,33
L-amino acid import (GO:0043092)	4,42E-02	1,00	22,22	69,33
regulation of t-circle formation (GO:1904429)	4,42E-02	1,00	22,22	69,33
positive regulation of aspartic-type endopeptidase activity involved in amyloid precursor protein catabolic process (GO:1902961)	4,42E-02	1,00	22,22	69,33
actin polymerization or depolymerization (GO:0008154)	2,67E-03	1,00	10,81	64,08
neuromuscular junction development (GO:0007528)	9,72E-03	1,00	13,33	61,79
positive regulation of receptor activity (GO:2000273)	2,88E-03	1,00	10,53	61,59
cell morphogenesis (GO:0000902)	1,07E-03	0,91	8,89	60,80
skeletal muscle myosin thick filament assembly (GO:0030241)	5,13E-02	1,00	19,05	56,56
positive regulation of insulin-like growth factor receptor signalling pathway (GO:0043568)	5,13E-02	1,00	19,05	56,56
membrane raft polarization (GO:0001766)	5,13E-02	1,00	19,05	56,56
positive regulation of keratinocyte proliferation (GO:0010838)	5,13E-02	1,00	19,05	56,56
regulation of high voltage-gated calcium channel activity (GO:1901841)	5,13E-02	1,00	19,05	56,56
regulation of myosin-light-chain-phosphatase activity (GO:0035507)	5,13E-02	1,00	19,05	56,56
positive regulation of glutamate receptor signalling pathway (GO:1900451)	5,13E-02	1,00	19,05	56,56
regulation of endocytic recycling (GO:2001135)	5,13E-02	1,00	19,05	56,56

negative regulation of axon regeneration (GO:0048681)	5,13E-02	1,00	19,05	56,56
left/right pattern formation (GO:0060972)	5,13E-02	1,00	19,05	56,56

Table 5: Top 50 miRNA associated with CUD host genes identified via Enrichr TargetScan 2017 database. Data output by Enrichr is listed according to Z-score as this is the most accurate measure, as described above (Chapter 2, page 20).

Term	P-value	Adjusted P-value	Odds Ratio	Z-Score
hsa-miR-3676	7,92E-03	6,76E-01	2,33	11,28
hsa-miR-1249	1,53E-02	6,52E-01	2,34	9,80
hsa-miR-4714-3p	7,67E-03	7,48E-01	1,69	8,21
hsa-miR-4497	2,42E-02	7,17E-01	2,06	7,67
hsa-miR-3669	4,54E-02	6,60E-01	2,33	7,19
hsa-miR-4774-5p	1,97E-02	7,10E-01	1,59	6,24
hsa-miR-3197	3,46E-02	8,14E-01	1,81	6,08
hsa-miR-523	1,41E-01	7,20E-01	3,03	5,93
hsa-miR-4523	5,75E-02	6,55E-01	2,04	5,82
hsa-miR-4732-5p	2,61E-02	6,87E-01	1,59	5,79
hsa-miR-1204	4,68E-02	6,53E-01	1,78	5,44
hsa-miR-1539	5,23E-02	6,61E-01	1,80	5,32
hsa-miR-4655-3p	5,64E-02	6,65E-01	1,85	5,31
hsa-miR-1180	7,07E-02	6,44E-01	1,94	5,14
hsa-miR-3131	6,51E-02	6,45E-01	1,87	5,12
hsa-miR-611	6,51E-02	6,35E-01	1,87	5,12
hsa-miR-4304	7,45E-02	6,44E-01	1,92	4,98
hsa-miR-3689b	3,89E-02	7,59E-01	1,50	4,88

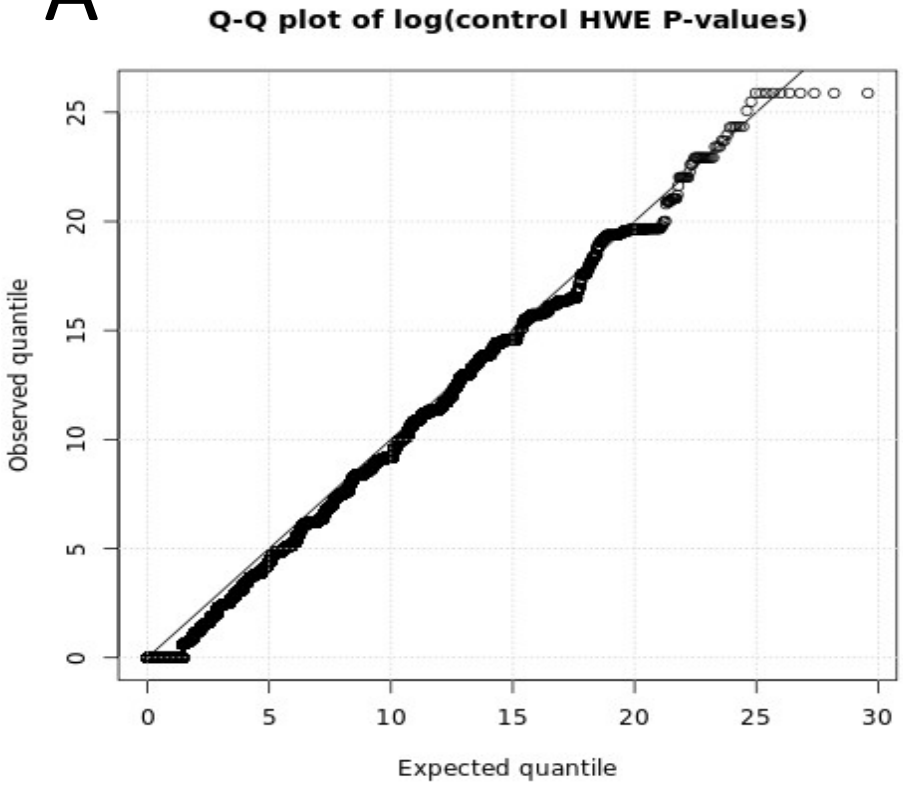
hsa-miR-3689a-5p	3,89E-02	7,38E-01	1,50	4,88
hsa-miR-3689e	3,89E-02	7,18E-01	1,50	4,88
hsa-miR-3689f	3,89E-02	6,99E-01	1,50	4,88
hsa-miR-3147	4,06E-02	6,30E-01	1,48	4,73
hsa-miR-3146	4,72E-02	6,44E-01	1,51	4,60
hsa-miR-1973	7,91E-02	6,35E-01	1,80	4,56
hsa-miR-4479	1,10E-01	6,71E-01	1,98	4,37
hsa-miR-3917	9,26E-02	6,59E-01	1,82	4,32
hsa-miR-718	1,04E-01	6,73E-01	1,86	4,21
hsa-miR-574-3p	1,06E-01	6,65E-01	1,86	4,16
hsa-miR-4671-3p	6,59E-02	6,34E-01	1,48	4,03
hsa-miR-887	1,12E-01	6,67E-01	1,83	3,99
hsa-miR-4781-5p	9,22E-02	6,63E-01	1,62	3,85
hsa-miR-4664-3p	1,20E-01	6,74E-01	1,79	3,79
hsa-miR-770-5p	7,37E-02	6,63E-01	1,41	3,67
hsa-miR-4712-5p	7,37E-02	6,54E-01	1,41	3,67
hsa-miR-4694-5p	7,79E-02	6,48E-01	1,40	3,57
hsa-miR-2964a-5p	8,96E-02	6,73E-01	1,39	3,35
hsa-miR-3201	9,00E-02	6,68E-01	1,39	3,34
hsa-miR-4791	9,00E-02	6,61E-01	1,39	3,34
hsa-miR-4431	1,61E-01	7,40E-01	1,74	3,18
hsa-miR-584	1,01E-01	6,90E-01	1,37	3,13
hsa-miR-4718	1,06E-01	6,69E-01	1,36	3,05
hsa-miR-451	1,72E-01	7,40E-01	1,70	2,99

hsa-miR-4508	1,27E-01	7,01E-01	1,45	2,98
hsa-miR-103b	1,18E-01	6,92E-01	1,38	2,96
hsa-miR-4633-5p	1,20E-01	6,80E-01	1,38	2,91
hsa-miR-3193	1,52E-01	7,26E-01	1,53	2,89
hsa-miR-190b	1,18E-01	6,84E-01	1,35	2,88
hsa-miR-190	1,18E-01	6,78E-01	1,35	2,88
hsa-miR-3124-5p	2,61E-01	7,51E-01	2,02	2,72
hsa-miR-3912	1,91E-01	7,22E-01	1,64	2,71

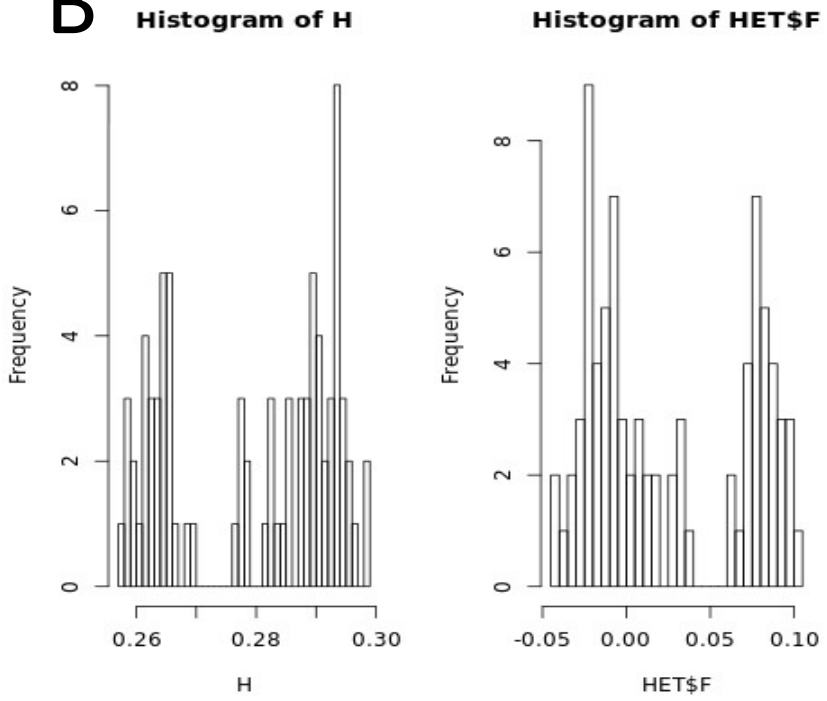
3.4 MUD GWAS quality control

All participants' data survived the quality control performed on the cohort. Of the 3946852 markers post imputation, 2653305 passed the filters used for quality control, with 1293524 SNPs being removed due to them being missing in a proportion of the subjects (Figure 5C) and 23 being removed due to deviations from HWE (Figure 5A); deviation from HWE could be due to genotyping errors, population stratification or true association (Turner *et al.*, 2011). Quality control visualisations were produced to confirm the presence of any poor-quality data needing to be removed (Figure 5 A-C). The quality control of the initial data (no imputation) yielded proportionally similar exclusions with 649422 of 654027 SNPs passing quality control with 19772 SNPs being removed due to them being missing in a proportion of the subjects (Supplementary Figure 1C), 198822 SNPs being removed due to not passing the MAF threshold and 16 being removed due to deviations from HWE (Supplementary Figure 1A). This indicated the effectiveness of the imputation with regards to increasing the amount of usable data in the study.

A



B



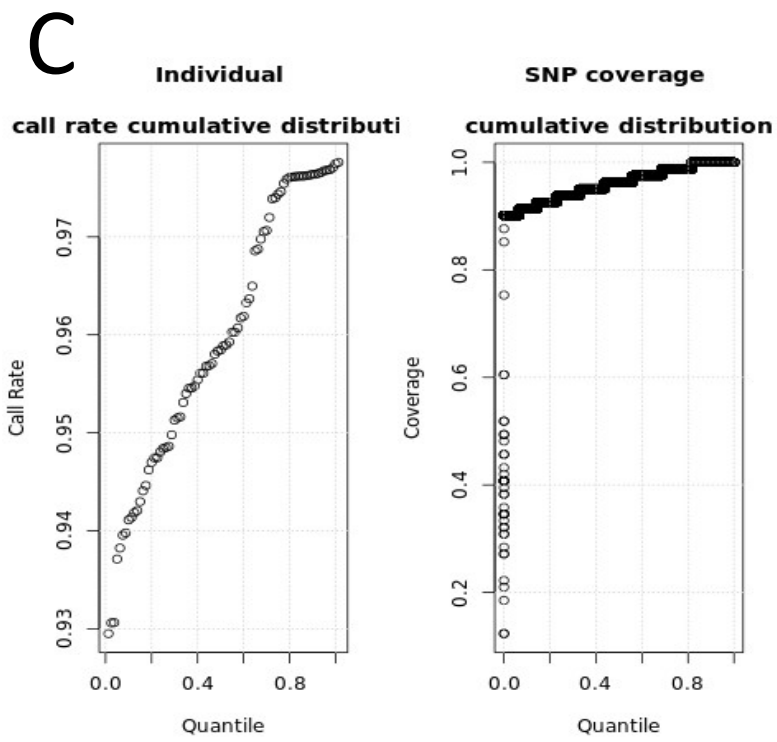


Figure 5: A) Q-Q plot of HWE indicating the expected vs observed values. The separation from the expected HWE values indicate SNPs deviating from the applied threshold which were removed following quality control. B) Histograms visualising the heterozygosity-based distributions measured either using F statistics (calculated by plink) or H (calculated using observed homozygotes and the number of non-missing autosomal genotypes). C) Quantile plots of genotype call rates and SNP coverage were used to confirm the viability of the data for further analyses. These indicate the SNPs passing the threshold of 95% (i.e. 95% of individuals in the study were not missing data for aforementioned SNPs), and the coverage of the SNPs read-depths in proportion to the quantiles of the theoretical cumulative distribution function with SNPs presenting low coverage being removed from the data after quality control.

3.5 MUD GWAS association analyses

The association analyses indicated that while no SNPs passed the genome-wide significance threshold of $5x10^{-8}$, 510 SNPs passed nominal GWAS significance threshold of $p<1x10^{-4}$ (Supplementary Table 3). These results were plotted using Manhattan (Figure 6) and Q-Q plots (Figure 7). The early separation of the observed values from the expected in Figure 7 is most likely attributed to population stratification, as expected due to the heterogeneity of the South African population. Genomic Inflation (Lambda (λ)) was calculated by Plink to be $\lambda = 2.08$. Lambda is calculated using the median observed

test statistic distribution per the expected test statistic distribution for the whole genome; this value is thus indicative of the deviation of the observed vs expected test statistic values with a higher value equating to greater inflation. Accepted lambda values generally fall around 1 with larger studies having values up to 1.42; as such accepted deviation depends on the size of the sample used in the study (Yang *et al.*, 2011).

The association analyses were also run with principle components (PCs) (n = 10) as covariates to adjust for the population stratification seen in Figure 7. Once the PCs were included as covariates the strength of association decreases globally and none of the SNPs passed the nominal or genome-wide significance thresholds of $1x10^{-4}$ and $5x10^{-8}$, respectively (Figure 8). Furthermore, although population stratification was adjusted for using the PCs and genomic inflation has decreased (λ. = 1.20), there is significant deflated deviation from the expected values due to sample size (Figure 9).

Genomic inflation can be affected by multiple factors such as population stratification, cryptic relatedness, and genotyping errors. These factors lead to inflated p-values represented by early inflated deviation from expected values, as seen in Figure 7 (Yang *et al.*, 2011).

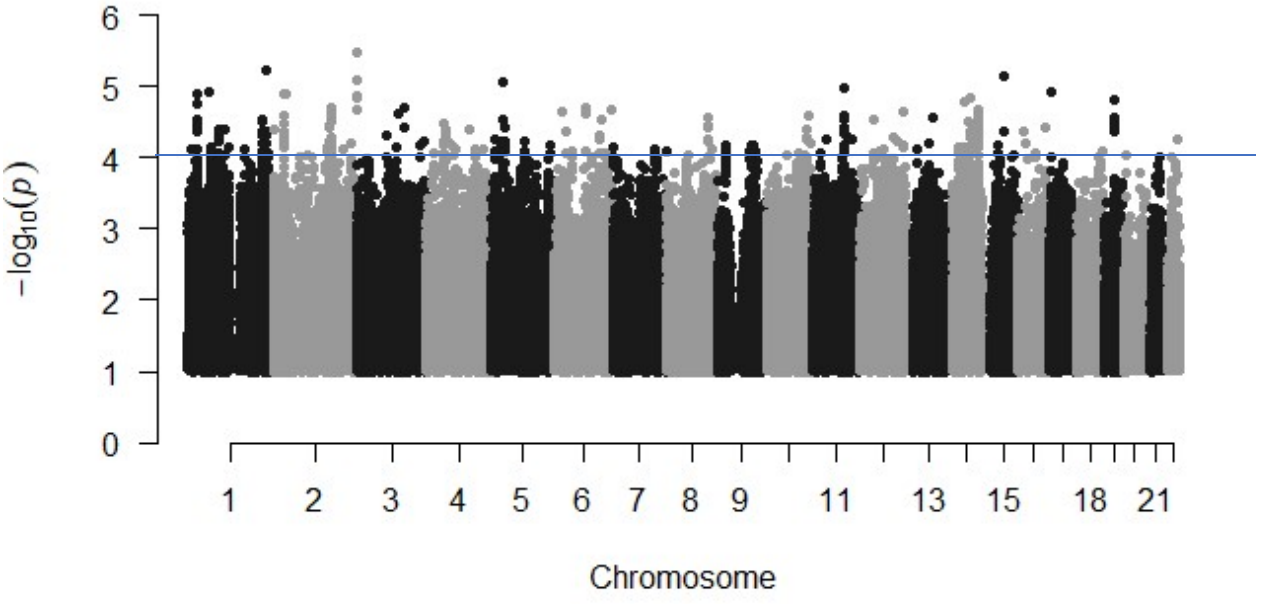


Figure 6: Manhattan plot of $-\log_{10}(P)$ logistic regression results for MUD cohort SNPs across each chromosome. A blue 'trend towards significance line' indicates SNPs passing the nominal significance threshold of $1x10^{-4}$ (n = 510).

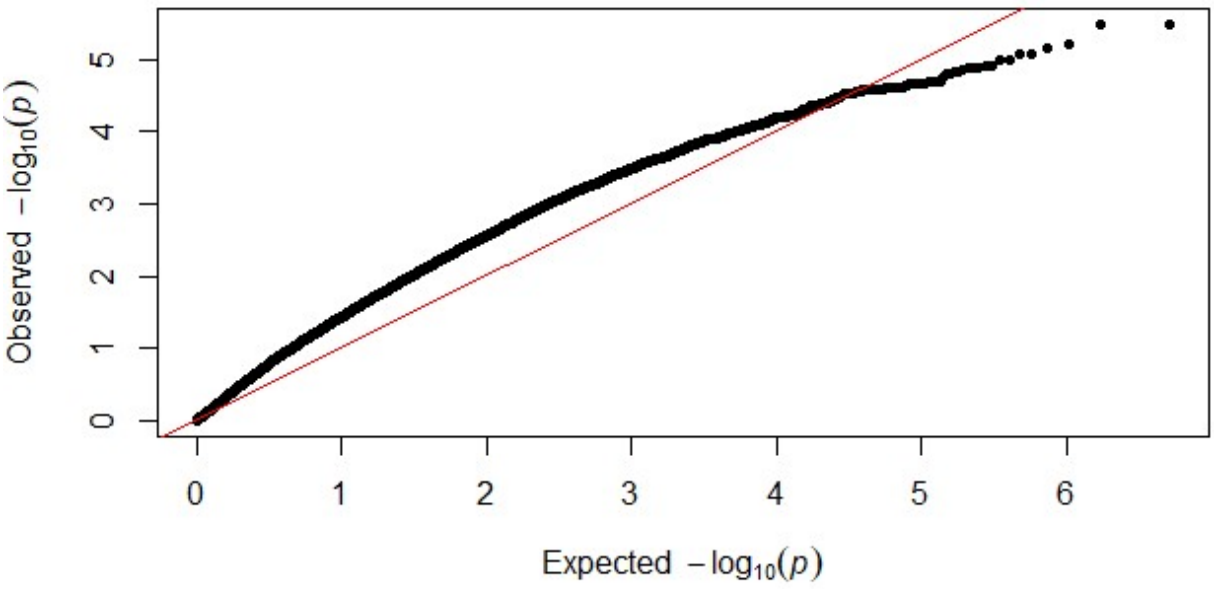


Figure 7: Q-Q plot of $-\log_{10}(P)$ results of logistic regression analysis of surviving MUD cohort SNPs plotted against their expected values. Lambda (λ) (genomic inflation) was calculated to be 2.08 by Plink.

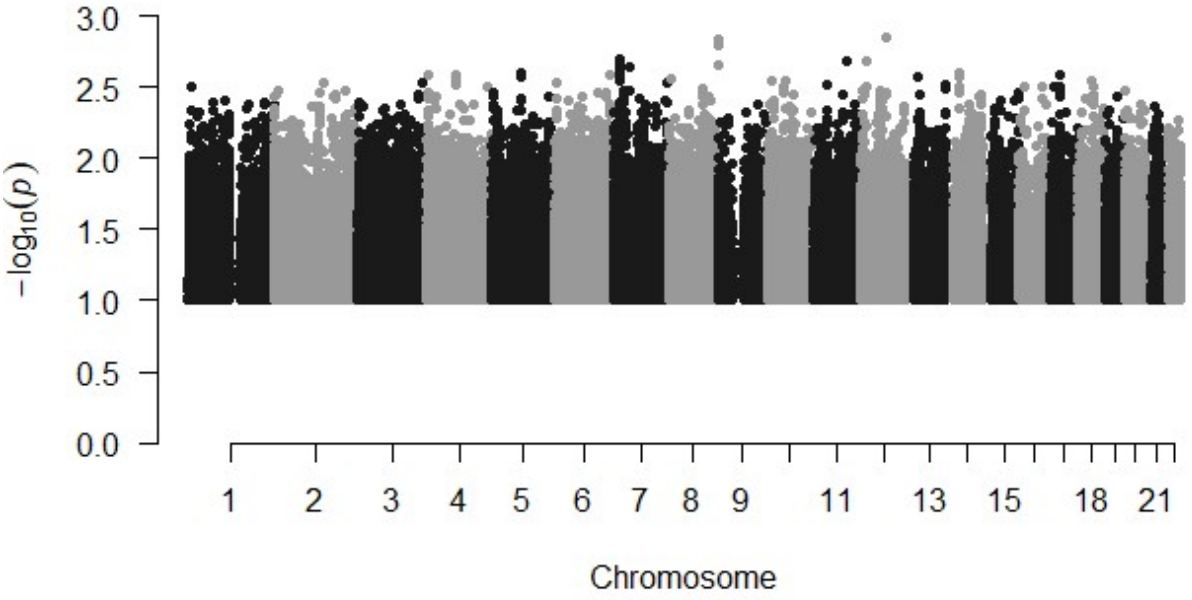


Figure 8: Manhattan plot of –log10(P) logistic regression results for MUD cohort SNPs across each chromosome after adjusting for population stratification using PCs. As expected, once population stratification is adjusted for, no SNPs pass either threshold.

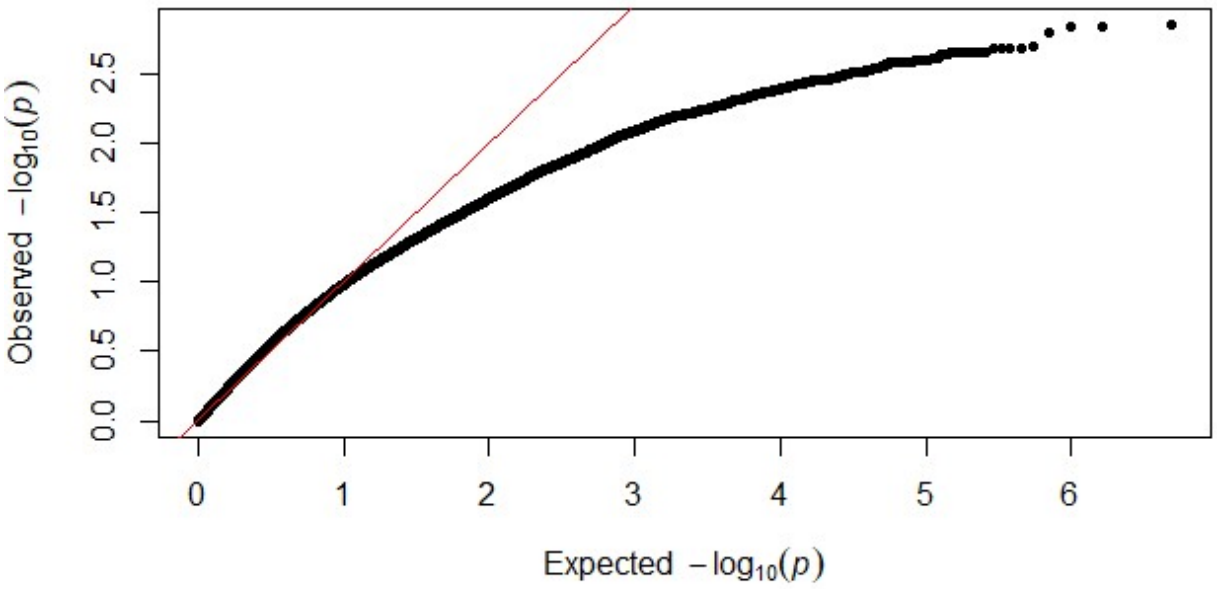


Figure 9: Q-Q plot of –log10(P) results of logistic regression analysis of surviving MUD cohort SNPs, after adjusting for population stratification using PCs, plotted against their expected values. Lambda (λ) (genomic inflation) was calculated to be 1.20 by Plink.

MUD GWAS results on the non-imputed data are provided in the supplement (Supplementary Figures 2-5).

3.6 MUD GWAS *in silico* analyses

Seeing that association analyses adjusting for PCs resulted in a loss of any SNPs approaching significance, the *in silico* analyses conducted for the MUD cohort all pertain to the results obtained without adjusting for PCs and as such do not account for population stratification in the MUD cohort. This means that the data used for *in silico* analyses are inflated; however as this is a pilot study, these results could become more reliable with increased cohort size should this become available (Price *et al.*, 2006; Ma and Amos, 2010; Shaukat, Rao and Khan, 2016). While PCA is the most appropriate method of accounting for population stratification and reduces inflation of the data, it does so at the cost of power to detect association which resulted in no detected associations; this does not necessarily equate to no associations, but simply a lack of association with the current cohort size.

Of the 510 SNPs identified by the association analyses (without PCA), only one presented any direct miRNA-mediated regulation potential when investigated using TarBase and SNPnexus; rs1805414 which was predicted to affect miR-29a-3p, miR-29b-3p, miR29c-3p and miR-21-3p. This is the largest change between the initial data (no imputation), as the initial data was unable to find any direct

miRNA-mediated regulation potential in the list of nominally significant SNPs. Enrichment analyses was performed using Enrichr yielding the top 50 GO biological processes for the 125 prioritised host genes identified by GAD (Supplementary Table 5). Investigation of miRNA-mediated regulation potential was performed for the 125 host genes identified via GAD, which were found to be regulated by several miRNA transcripts (Supplementary Table 6). The enrichment analyses performed using the initial data (no imputation) was only done using 116 identified using GAD; there are a number of different genes between the initial and imputed dataset however the biological processes identified remained consistent, with only the miRNA predicted via enrichment changing significantly (Supplementary Tables 8 and 9).

3.7 Cohort comparison results

The host genes of the MUD cohort were then compared to those identified in the CUD cohort, yielding a list of 57 genes that were exclusively identified in the MUD cohort (Supplementary Table 7). These host genes then underwent separate enrichment analyses, via Enrichr yielding the top 50 GO biological processes (Table 6) and miRNA (Table 7) associated with the genes associated exclusively with the MUD data. This was also done for the initial dataset which yielded 56 MUD unique genes (Supplementary Table 10), with enrichment analyses again outputting similar biological processes but with some miRNA predictions differing (Supplementary Tables 11 and 12).

Table 6: Top 50 (5%) GO biological processes associated with MUD exclusive host genes identified using Enrichr. Biological processes output are ranked by Z-score as this is the most accurate measure as described above (Chapter 2, page 20).

Term	P-value	Adjusted P-value	Odds Ratio	Z-Score
positive regulation of ion transmembrane transport (GO:0034767)	2,94 E-04	0,75	76,63	623,22
positive regulation of sodium ion transmembrane transporter activity (GO:2000651)	4,47 E-04	0,76	62,70	483,57
positive regulation of muscle hypertrophy (GO:0014742)	7,35 E-04	0,94	49,26	355,42
positive regulation of transporter activity (GO:0032411)	7,35 E-04	0,75	49,26	355,42
positive regulation of cardiac muscle hypertrophy (GO:0010613)	9,66 E-04	0,82	43,10	299,23
hyaluronan catabolic process (GO:0030214)	9,66 E-04	0,70	43,10	299,23
positive regulation of dendrite extension (GO:1903861)	1,09 E-03	0,70	40,57	276,62

regulation of dendrite extension (GO:1903859)	1,23 E-03	0,70	38,31	256, 81
vesicle-mediated transport in synapse (GO:0099003)	1,54 E-04	0,79	28,74	252, 19
regulation of mesenchymal stem cell differentiation (GO:2000739)	1,73 E-02	1,00	57,47	233, 24
regulation of oxidative stress-induced neuron intrinsic apoptotic signalling pathway (GO:1903376)	1,73 E-02	1,00	57,47	233, 24
regulation of skeletal muscle satellite cell proliferation (GO:0014842)	1,73 E-02	1,00	57,47	233, 24
negative regulation of interleukin-6 secretion (GO:1900165)	1,73 E-02	1,00	57,47	233, 24
cellular triglyceride homeostasis (GO:0035356)	1,73 E-02	1,00	57,47	233, 24
regulation of NAD(P)H oxidase activity (GO:0033860)	1,73 E-02	1,00	57,47	233, 24
regulation of adiponectin secretion (GO:0070163)	1,73 E-02	1,00	57,47	233, 24
positive regulation of sodium ion transmembrane transport (GO:1902307)	1,73 E-02	1,00	57,47	233, 24
negative regulation of amino acid transport (GO:0051956)	1,73 E-02	1,00	57,47	233, 24
regulation of cardiac muscle hypertrophy (GO:0010611)	1,68 E-03	0,85	32,84	209, 91
hyaluronan metabolic process (GO:0030212)	1,68 E-03	0,78	32,84	209, 91
regulation of monocyte chemotactic protein-1 production (GO:0071637)	2,01 E-02	1,00	49,26	192, 40
positive regulation of myeloid cell apoptotic process (GO:0033034)	2,01 E-02	1,00	49,26	192, 40
positive regulation of calcineurin-NFAT signalling cascade (GO:0070886)	2,01 E-02	1,00	49,26	192, 40
mitochondrial DNA metabolic process (GO:0032042)	2,01 E-02	1,00	49,26	192, 40
regulation of cardiac muscle cell membrane potential (GO:0086036)	2,01 E-02	1,00	49,26	192, 40
myoblast fusion (GO:0007520)	2,01 E-02	1,00	49,26	192, 40
regulation of endothelial cell chemotaxis to fibroblast growth factor (GO:2000544)	2,01 E-02	1,00	49,26	192, 40
malate metabolic process (GO:0006108)	2,01 E-02	1,00	49,26	192, 40
positive regulation of autophagy of mitochondrion (GO:1903599)	2,01 E-02	1,00	49,26	192, 40
positive regulation of ion transmembrane transporter activity (GO:0032414)	2,01 E-03	0,86	29,99	186, 18
peptidyl-cysteine modification (GO:0018198)	2,30 E-02	1,00	43,10	162, 65

positive regulation of protein homodimerization activity (GO:0090073)	2,30 E-02	1,00	43,10	162,65
response to epinephrine (GO:0071871)	2,30 E-02	1,00	43,10	162,65
phosphatidylinositol-3-phosphate biosynthetic process (GO:0036092)	2,30 E-02	1,00	43,10	162,65
glycerol-3-phosphate metabolic process (GO:0006072)	2,30 E-02	1,00	43,10	162,65
negative regulation of anion transmembrane transport (GO:1903960)	2,30 E-02	1,00	43,10	162,65
lagging strand elongation (GO:0006273)	2,30 E-02	1,00	43,10	162,65
negative regulation of potassium ion transmembrane transporter activity (GO:1901017)	2,30 E-02	1,00	43,10	162,65
cellular response to epinephrine stimulus (GO:0071872)	2,30 E-02	1,00	43,10	162,65
negative regulation of biosynthetic process (GO:0009890)	2,30 E-02	1,00	43,10	162,65
regulation of ATP biosynthetic process (GO:2001169)	2,30 E-02	1,00	43,10	162,65
polyol biosynthetic process (GO:0046173)	2,30 E-02	1,00	43,10	162,65
carboxylic acid catabolic process (GO:0046395)	2,98 E-03	1,00	24,63	143,28
positive regulation of leukocyte apoptotic process (GO:2000108)	2,58 E-02	1,00	38,31	140,12
negative regulation of gluconeogenesis (GO:0045721)	2,58 E-02	1,00	38,31	140,12
response to muscle stretch (GO:0035994)	2,58 E-02	1,00	38,31	140,12
negative regulation of protein sumoylation (GO:0033234)	2,58 E-02	1,00	38,31	140,12
regulation of protein acetylation (GO:1901983)	2,58 E-02	1,00	38,31	140,12
negative regulation of response to reactive oxygen species (GO:1901032)	2,58 E-02	1,00	38,31	140,12
positive regulation of vascular endothelial cell proliferation (GO:1905564)	2,58 E-02	1,00	38,31	140,12

Table 7: Top 50 miRNA associated with MUD exclusive host genes based off enrichment analysis. Predicted miRNA are ranked by Z-score as this is the most accurate measure as described above (Chapter 2, page 20).

Term	P-value	Adjusted P-value	Odds Ratio	Z-Score
hsa-miR-382	0,02	1,00	9,58	38,19

hsa-miR-335	0,03	1,00	7,93	28,81
hsa-miR-223	0,03	1,00	7,04	24,03
hsa-miR-202	0,04	1,00	6,76	22,59
hsa-miR-26A, hsa-miR-26B	0,01	1,00	4,58	20,55
hsa-miR-22	0,03	1,00	4,46	15,67
hsa-miR-507	0,06	1,00	5,26	15,22
hsa-miR-324-5P	0,14	1,00	6,90	13,79
hsa-miR-383	0,14	1,00	6,90	13,79
hsa-miR-200A	0,15	1,00	6,39	12,32
hsa-miR-486	0,16	1,00	5,95	11,08
hsa-miR-526C, hsa-miR-518F, hsa-miR-526A	0,16	1,00	5,95	11,08
hsa-miR-199A, hsa-miR-199B	0,08	1,00	4,31	10,96
hsa-miR-30A-5P, hsa-miR-30C, hsa-miR-30D, hsa-miR-30B, hsa-miR-30E-5P	0,03	1,00	2,96	10,75
hsa-miR-323	0,08	1,00	4,23	10,62
hsa-miR-345	0,16	1,00	5,75	10,53
hsa-miR-432	0,16	1,00	5,65	10,27
hsa-miR-105	0,09	1,00	3,92	9,33
hsa-miR-25, hsa-miR-32, hsa-miR-92, hsa-miR-363, hsa-miR-367	0,06	1,00	3,33	9,28
hsa-miR-202	0,10	1,00	3,85	9,06
hsa-miR-508	0,18	1,00	5,15	8,91
hsa-miR-302B	0,18	1,00	5,00	8,52
hsa-miR-122A	0,20	1,00	4,60	7,49
hsa-miR-296	0,20	1,00	4,60	7,49
hsa-miR-19A, hsa-miR-19B	0,06	1,00	2,66	7,35
hsa-miR-30A-3P, hsa-miR-30E-3P	0,12	1,00	3,40	7,28
hsa-miR-499	0,20	1,00	4,48	7,19
hsa-miR-381	0,12	1,00	3,35	7,09
hsa-miR-136	0,21	1,00	4,31	6,77
hsa-miR-518A-2	0,13	1,00	3,24	6,68
hsa-miR-34B	0,13	1,00	3,12	6,26
hsa-miR-103, hsa-miR-107	0,13	1,00	3,12	6,26
hsa-miR-218	0,11	1,00	2,57	5,65
hsa-miR-150	0,23	1,00	3,83	5,62
hsa-miR-130A, hsa-miR-301, hsa-miR-130B	0,11	1,00	2,55	5,57
hsa-miR-9	0,15	1,00	2,92	5,55
hsa-miR-361	0,24	1,00	3,71	5,34
hsa-miR-302C	0,16	1,00	2,84	5,26
hsa-miR-324-3P	0,25	1,00	3,55	4,99
hsa-miR-500	0,25	1,00	3,48	4,82
hsa-miR-505	0,26	1,00	3,28	4,38
hsa-miR-484	0,27	1,00	3,19	4,18
hsa-miR-200B, hsa-miR-200C, hsa-miR-429	0,16	1,00	2,20	4,10
hsa-miR-452	0,28	1,00	3,08	3,94

hsa-miR-374	0,20	1,00	2,42	3,89
hsa-miR-217	0,28	1,00	3,00	3,77
hsa-miR-9	0,18	1,00	2,06	3,57
hsa-miR-506	0,16	1,00	1,89	3,44
hsa-miR-96	0,22	1,00	2,27	3,43
hsa-miR-516-3P	0,30	1,00	2,78	3,32

Chapter 4 Discussion

This study aimed to further elucidate the potential regulatory role of miRNA in MUD by using a PGC-CUD discovery cohort alongside a local MUD cohort. The study findings indicate that while individuals with CUD or MUD are associated with similar genes, there are also several genes unique to the MUD cohort that have been previously associated with MUD aetiology, such as *PARP1*, *UNC13A*, *PARK2, PARK7* and several *SLC* family genes. Subsequently, analyses predicted regulation by several miRNA that have been previously associated with MUD as well as other SUDs, such as miR-9, miR-21, miR-22, miR-29a, miR-29b, miR-29c, miR-96 and miR-200. Furthermore, the enrichment analyses identified several biological processes associated with these genes that were involved in MUD aetiology, such as processes involved in synaptic transport mechanisms, autoinflammatory immune response, autophagy, apoptosis, and synaptic plasticity.

In this way, the study further elucidated the molecular underpinnings of MUD with regards to miRNA-mediated regulation. The study identified miR-let-7e-5p and miR-140-5p to be directly involved in CUD, with subsequent enrichment analyses also indicating association of several other miRNA, such as miR-190b, miR-196a-5p, miR-196b-5p and miR-210-3p, via the CUD discovery cohort. Studies have previously shown that these miRNA are associated with MUD as well as the progression of MUD through regulation of processes related to MA neurotoxicity and synaptic plasticity (Dreyer, 2010; R. Wei *et al.*, 2016; Zhao *et al.*, 2016). Direct miRNA-mediated regulation was found in one MUD nominally associated SNP, rs1805414, which was predicted to be regulated by miR-21-3p and several miR-29 family miRNA. Furthermore, enrichment analyses identified several miRNA statistically associated with MUD including miR-182, miR-125 and miR-200 family miRNA, all of which have been previously associated with MUD disorder aetiology (Zhao *et al.*, 2016; Sim *et al.*, 2017; Li *et al.*, 2018). Further enrichment predictions were made based off a gene list comparison between the two cohorts, MUD and CUD, used in this study. This comparison identified several genes that were associated with both CUD and MUD, as well as several genes that were unique to the MUD cohort, such as *PARP1*, *UNC13A*, *PARK2*, *PARK7* and several *SLC* family genes. These enrichment analyses yielded several biological pathways previously associated with substance dependence and MUD aetiology, such as apoptosis, autophagy, and synaptic transport mechanisms, in which these genes are involved. The enrichment analyses also identified several miRNA involved in MUD aetiology to be related to these MUD exclusive genes such as miR-9, miR-22, miR-96, and miR-200.

4.1 CUD as a discovery cohort

Fifty-seven genes were found to be nominally associated with both MUD and CUD. All the genes identified by the *in silico* bioinformatic pipeline were used in an enrichment analysis that identified several miRNA (miR-190, miR-190b, let-7e-5p, miR-196a-5p, miR-196b-5p, miR-140-5p and miR-210-3p) and biological processes related to stimulant abuse/dependence, such as synaptic transport mechanisms (GO:0060402, GO:2001259, GO:1901021, GO:0010359, GO:0071286), autophagy and apoptosis (GO:0001773, GO:0050855, GO:0002664) (Cao et al., 2016; Guha, Harraz and Snyder, 2016). This is supported by studies investigating cocaine neurotoxicity and cellular damage, which found that cocaine's neurotoxicity causes autophagic cell death via a signalling cascade that utilizes nitric oxide (Guha, Harraz and Snyder, 2016) and is also associated with apoptosis (Cao *et al.*, 2016). This is further supported by studies investigating these miRNA. One study found that miR-196a-5p targets genes involved in neuronal differentiation and cell death such as *ANXA1*, a mediator of apoptosis (Kunkanjanawan *et al.*, 2016), while another study in rats found that miR-210-3p was capable of affecting the regulation of inflammation and apoptosis through its effects on *SERPINE1*, a gene involved in several inflammatory pathways (Zhang *et al.*, 2019). Other studies investigating miR-let-7e-5p and miR-190 found that miR-let-7e-5p had an inhibitory effect on apoptosis while miR-190 displays an inhibitory effect on neuroinflammation (Zhang *et al.*, 2018; Q. Sun *et al.*, 2019). This indicates the role of miRNA in the biological processes associated with CUD via the enrichment analyses and thus their subsequent role on cocaine-induced neurotoxicity. This is echoed in similar studies speaking to neurotoxicity in MUD, with one such study indicating that miRNA-mediated regulation of *BBC3*, a gene regulating the relationship between autophagy and apoptosis, as a result of MA use could regulate microglial survival and thus mediate the role of autophagy and apoptosis in MA-induced neurotoxicity (Zhang *et al.*, 2016). Another study found that MA induced increased expression of apoptotic genes, such as *BBC3* and *P53*, and also identified differential expression of autophagy-related genes such as, *ATG12* and *MAP1LC3B*, with *MAP1LC3B* having pro-apoptotic effects; the study stipulates that MA use induces neurotoxicity-associated signalling cascades involved in apoptosis and autophagy (Tehrani *et al.*, 2019). With studies already investigating the effects of miRNA-mediated regulation on genes involved in MA neurotoxicity (Zhang *et al.*, 2016) it would not be surprising for studies to focus more on the specific roles of miRNA in MA-induced autophagy and apoptosis.

The comparison of MUD to CUD identified 57 genes that are exclusively related to nominally significant SNPs (see section 2.7 and Supplementary Table 7) of the MUD association analysis. Enrichment of these genes identified several biological processes associated with MUD and MA

neurotoxicity, such as synaptic transport mechanisms (GO:0034767, GO:0099003 and GO:1902307), autoinflammatory immune response (GO:1900165), apoptosis (GO:1903376, GO:0033034 and GO:2000108) and autophagy (GO:1903599) and synaptic plasticity (GO:1903861 and GO:1903859). The enrichment of these genes also yielded a miRNA regulatory network comprised of miRNA known to regulate processes involved in MUD aetiology as well as miRNA previously associated with other SUDs (Buchanan, Sparkman and Johnson, 2010; Zhang et al., 2016; Huang et al., 2017; Kays and Yamamoto, 2019; Sun et al., 2019). Several studies have already been conducted investigating the role of miRNA in these processes, with one study identifying differential expression of miRNA involved in synaptic plasticity and synaptic transmission as a result of MA exposure (Sim *et al.*, 2017). Another study investigating miRNA expression in MA-exposed mice suggested that miRNA being regulated by MA exposure contribute to neuronal autophagy and immune response (Zhu *et al.*, 2015); this is further supported by other studies investigating the role of miRNA in drug abuse where differential expression was identified in miRNA associated with adaptive immunity and neurotransmitter release in dopamine neurons (Zhang, Jing and Wang, 2016).

4.2 MUD association and *in silico* analyses

4.2.1 Genetic association analyses

While no SNPs passed the genome-wide significance threshold in the MUD cohort, 125 SNPs passed the nominal significance threshold. Only one of the surviving SNPs were directly involved in miRNA-mediated regulation, however, the genes wherein the surviving SNPs were located had miRNA-regulatory potential. This SNP, rs1805414, mapped to the *PARP1* gene which is known to be involved in neurodegenerative diseases and plays a role in regulating apoptosis (Martire, Mosca and d'Erme, 2015). This is further validated in a rat study that found that methamphetamine induced an increase in PARP1 protein expression indicating apoptosis activation (Wang *et al.*, 2017). The SNP indicated direct involvement with four miRNA, miR-21-3p, miR-29a-3p, miR-29b-3p and miR29c-3p. Of these miRNA, a study by Zhu *et al.* (2015), indicated differential expression of miR-29b and miR-29c as a result of MA exposure and suggested that miRNA contributed to MA addiction through regulation of the reward pathways affected by MA as well as regulation of synaptic plasticity; with other studies also indicating a role in apoptosis (Liu *et al.*, 2016; Chen *et al.*, 2017). Another study found that miR-21-3p was able to regulate cellular apoptosis and inflammation through its regulation of *MAT2B* (Ge *et al.*, 2018), with both pathways known to be involved in MUD aetiology. A recent study investigating the effects of Rhynchophylline, a medicinal herb, on conditioned place preference (CPP) rats conditioned by methamphetamine found the differential expression of miR-21 in serum exosomes (Li *et al.*, 2018) with miR-21 known to be expressed in dendritic cell subsets

while also inhibiting many pro-inflammatory cytokine production pathways and offering protection against apoptosis (Smyth *et al.*, 2015; Han *et al.*, 2017; He *et al.*, 2017; Yu *et al.*, 2019). miR-29a was also found to be involved in apoptosis by another study which found that knock-down of miR-29a resulted in an upregulation of *P21* and *P53* which are known to be involved in apoptosis induction (Khamisipour *et al.*, 2018). These studies indicate that this SNP is involved direct miRNA-mediated regulation by several miRNA that are involved in processes known to be involved in MUD aetiology (Zhang *et al.*, 2016; Tehrani *et al.*, 2019). This association is also found in the subsequent enrichment analyses of the host genes identified. The genes identified included several of the genes identified in the CUD cohort summary statistics as well as several others unique to the MUD cohort data; such as *UNC13A* a key component to synaptic transport mechanisms due to its role in vesicle fusion at defined release sites (Reddy-Alla *et al.*, 2017). Another gene identified here, i.e. *PARK2,* is associated with autophagy (Wang *et al.*, 2012), along with *PARK7* (B. Wang *et al.*, 2016) which can protect against MA neurotoxicity (Moszczynska, 2016).

4.2.2 Enrichment analyses

Gene enrichment analyses revealed several biological processes associated with synaptic plasticity, such as dendritic spine morphogenesis and regulation of dendrite extension (GO:0060997 and GO:1903859), and immune response, such as negative regulation of interleukin-6 secretion and regulation of cellular pH (GO:1900165 and GO:0030641) (Lippman and Dunaevsky, 2005; Esnafoglu and Ayyıldız, 2017; Buechler and Turley, 2018; Erra Díaz, Dantas and Geffner, 2018). The enrichment analyses also identified miRNA targeting these genes, several of which were found to be previously associated with other SUDs, such as miR-125, miR-182 and miR-200 which have been associated with CUD (Eipper-Mains *et al.*, 2011) and miR-9 which has been previously associated with alcohol use disorder (Treistman and Martin, 2009). While there were no major differences to the biological processes predicted via enrichment of the initial dataset (no imputation), there were a number of predicted miRNA that differed with the initial dataset identifying other miRNA associated with MUD and related processes such as miR-21 and miR-210 which were previously associated with other SUDs such as alcohol use disorder and nicotine dependence.

When the MUD unique genes were separately assessed via enrichment for different biological processes, gene ontologies and miRNA, the biological processes associated were involved in immune response, dendritic spine morphogenesis and synaptic transport mechanisms. This is not surprising as MA has been shown to decrease the amount of dendritic cells and has shown to produce an immunosuppressive effect (Harms *et al.*, 2012), it has also been shown that MA produces abnormal changes in neurotransmission and neuronal plasticity due to its effects on synaptic transport

mechanisms such as the voltage-gated calcium channels (Cadet, Bisagno and Milroy, 2014; González *et al.*, 2016). Studies have shown that drug-induced changes to dendritic spine density and plasticity are important factors for addictive behaviours (Hyman, Malenka and Nestler, 2006), with another study investigating whether MA-induced effects on dendritic spine structure were resultant of conditioning or not indicated that the increase in spine density was related to drug-associated memory acquisition (Singer *et al.*, 2009); in short, MA induces several changes to dendritic spine density and plasticity subsequently affecting drug-associated learning and addictive behaviours (Zhao *et al.*, 2019). Methamphetamine is known to affect various facets of the immune system ranging from innate and adaptive immune response to dendritic cells (Mahajan *et al.*, 2006; Martinez *et al.*, 2009; Sriram *et al.*, 2016). A recent study investigating the effect of MA on T cell proliferation found that MA induced differential regulation of genes involved in cell cycle progression and suggested that this dysregulation could result in inhibited T cell proliferation thus negatively effecting immune response (Potula *et al.*, 2018). Thus, enrichment analyses of the MUD associated genes were able to identify several biological processes involved in MUD disorder aetiology.

4.2.3 Investigating the role of miRNA

The enrichment of these MUD unique genes also identified several miRNA that were also found to regulate genes in other SUDs as well as MUD and were previously associated with SUDs such as CUD and alcohol use disorder (O'Connor *et al*., 2016). These include miR-96 and miR-200 which have been previously associated with apoptosis and autophagy, and miR-9 and miR-22 which has been previously associated with synaptic plasticity and dendritic extension. miR-9 has been shown to interact with *RAP2A*, a gene which is associated with synaptic plasticity as well as dendritic spine morphology (Stornetta and Zhu, 2011; Sim *et al.*, 2016; Xue *et al.*, 2016) but is also shown to be associated with apoptosis through its effects on *BCL2L11*, a gene known to be associated with toxicity induced autophagy and apoptosis (Luo and Rubinsztein, 2013; N. Wei *et al.*, 2016). Likewise, miR-22 is also shown to be involved in synaptic plasticity, through its effects on *CBEB* which is involved in long-term memory related synaptic plasticity (Fiumara *et al.*, 2015; Hu and Li, 2017), as well as apoptosis through its interaction with several apoptosis-related genes such as *CASP3*, *BCL2* and *eIF4EBP3* (Ma *et al.*, 2016; Lv *et al.*, 2018). miR-200 was shown to be involved in oxidative-stress induced apoptosis and simulated neurotoxin induced apoptosis through an interaction with *P53*, a known apoptosis related gene (Pandey *et al.*, 2015; Wei *et al.*, 2015). Similarly, miR-96 was shown to be associated with apoptosis and was found to inhibit autophagy through an interaction with *FOXO1* which has been associated with both autophagy and apoptosis (Hariharan Nirmala *et al.*, 2010; Haflidadóttir *et al.*, 2013; Shi *et al.*, 2017). The enrichment of MUD unique genes from the

initial non-imputed dataset also yielded several other predicted miRNA such as miR-142-3p which has been previously associated with dendritic cell function and immune response (Smyth *et al.*, 2015), miR-126 and miR-516a family miRNA, which have been previously associated with drug metabolism (He *et al.*, 2015). miR-142-3p is predicted to interact with interleukin 6 (IL-6) with studies having shown that MA use significantly upregulates IL-6 secretion (Smyth *et al.*, 2015; Papageorgiou *et al.*, 2019). miR-126 has been shown to interact with *SLC* family genes (He *et al.*, 2015), many of which are involved in synaptic transport mechanisms (Hu *et al.*, 2020), while miR-516a interacts with *SULF1* a gene known to be associated to drug metabolism (He *et al.*, 2015). A recent study, investigating the effects of miR-142-3p on a neuroinflammatory gene *PELI1*, indicated that miR-142-3p produced an inhibitory effect towards MA-induced neuroinflammation by targeting *PELI1* and stipulated that modifications to the miR-142-3p/*PELI1* axis could be a potential therapeutic strategy for MA-mediated neuroinflammation (Yu *et al.*, 2019). These and many other miRNA were also found to be associated with MUD aetiology through proxy studies; a rat study indicated differential expression of miR-200 and miR-9 as a result of MA exposure (Sim *et al.*, 2017) with another study indicating differential expression of miR-22 as a result of MA exposure in serum-exosomes (Zhao *et al.*, 2016). These studies indicate the role that miRNA identified in this study have with regards to MUD aetiology through their involvement in several pathways associated with MUD.

From these findings it can be inferred that while SUDs share common molecular mechanisms and pathways, there may be regulatory differences amongst the different SUDs. This emphasizes the importance of understanding the miRNA networks influencing these pathways and mechanisms to properly elucidate the specific disorder's aetiology.

4.3 Limitations

4.3.1 Cohort size

Several limitations need to be considered. The relatively small sample size of the MUD cohort likely resulted in the lack of SNPs passing genome-wide significance. However, as this is an exploratory study that makes use of a large CUD discovery cohort to assist identification of relevant data, the sample size is suitable for use in this manner.

4.3.2 *In silico* online tools

The online tools used to perform *in silico* analyses and enrichment analyses often use third-party databases, meaning that not all current research is able to be referenced by the database when accessed, as some of these databases are no longer being curated. Even when the databases are up to date there

is very little literature and information available on MUD and as a result the database cannot identify strong associations between the genes and the MUD phenotype.

4.3.3 Diagnostic criterion and separate study sites

The patients were all diagnosed for a separate study by Lanesman *et al* (2019) and as such, many of the samples that were collected and used for DNA extraction were done using different protocols leading to extraction efficiency differences and as such some subjects samples were unable to be assayed and used in the study and could not be included in the study cohort. Furthermore, the cohort used consisted of South African mixed-ancestry individuals and as such consisted of a very heterogeneous population, which could not be accounted for in the association analyses due to sample size.

4.3.4 Covariate and confounding data not accounted for

Population stratification due to the mixed ancestry of the cohort could not be accounted for due to the small sample size. As seen in Figure 7, there was a high level of genomic inflation in this study as is evidenced by the inflation of the observed test statistics seen deviating from the expected values as the population stratification being left unaccounted for results in artificial differences in allele frequencies affecting the investigated SNPs and subsequently the test statistics.

As mentioned in Materials and Methods (Chapter 2), population stratification was accounted for by way of PCA, however as this resulted in further reducing the power to detect association in this study and left no SNPs passing the genome-wide or nominal significance thresholds ($5x10^{-8}$ and $1x10^{-4}$, respectively) this was subsequently left unaccounted for, resulting in the SNPs passing the nominal significance threshold of $1x10^{-4}$ but at the cost of a high level of genomic inflation which could indicate false positives.

Nonetheless, this study was able to identify several miRNA and target genes trending towards significant association with MUD disorder aetiology. These findings, discussed above, are consistent with the little MUD literature currently available as well as literature on other SUDs. Further investigation of the underlying MUD-mediated regulatory network is warranted.

4.4 Future considerations

In future, studies should investigate MUD as compared to several other SUDs as opposed to just one (CUD) to identify SNPs and genes common to all, as well as genes unique to MUD. As with any NP disorder study, cross-disorder studies should be conducted to understand the similarities between MUD and other NP disorders. Studies should also consider other epigenetic regulatory networks, such as DNA methylation, involved in the regulation of the disorder. Studies should also aim to

investigate these factors in diverse population groups as these groups have the potential to produce more diverse information than homogeneous population groups typically studied. Additionally, studies should account for population stratification (to avoid false-positives) and should consider using imputation with newer South African-centric reference genomes, where applicable, to enhance the ability to detect associations with MUD. Finally, functional validation of putative miRNA mechanisms should be performed to validate the effect of miRNA-mediated regulation in MUD and MUD related biological processes and pathways. Hopefully, in the future, more work will be done on this disorder in a local context to enhance and/or broaden the knowledge scope of MUD to alleviate the socio-economic burden caused by the disorder.

4.5 Conclusion

In this study a CUD discovery cohort was used alongside association analyses of a local cohort with MUD and *in silico* online tools to investigate the role of miRNA-mediated regulation in MUD. The objectives of the study were to perform a logistic regression association analysis of the MUD cohort, and use bioinformatics and enrichment analyses on both the CUD and MUD cohorts before comparing the bioinformatic profiles of each cohort to elucidate the role of miRNA in MUD.

To summarize, this study indicates an epigenetic component to MUD aetiology through miRNA-mediated regulation of genes involved in biological processes related to MUD pathophysiology. This was accomplished despite using a relatively small mixed-ancestry cohort, with high levels of genomic admixture affecting population stratification, where population-stratification was left unaccounted for leading to high levels of genomic inflation. Thus, it is plausible that with a larger sample size and fewer confounding factors (like population stratification), these results may increase in significance or allow for more accurate results.

In conclusion, this study contributes to knowledge on the molecular underpinnings of MUD, highlighting the differences between MUD and other stimulant use disorders like CUD. It also suggests an epigenetic component to MUD aetiology in the form of miRNA-mediated regulation. As such, the need for better characterisation of these disorders is indicated; subsequently leading to more objective ways of diagnosis. This study contributes to the understanding of how miRNA regulatory networks function in MUD through their interaction with genes both exclusive to MUD and common amongst this and other SUDs. Furthermore, this is the first study investigating the molecular underpinnings of MUD in a South African cohort, indicating the potential for use of local mixed-ancestry populations to identify novel variants associated with miRNA-mediated regulation in MUD disorder aetiology.

Chapter 5 References

Alural, B., Genc, S. and Haggarty, S. J. (2017) 'Diagnostic and therapeutic potential of microRNAs in neuropsychiatric disorders: Past, present, and future', *Progress in neuro-psychopharmacology & biological psychiatry*, 73, pp. 87–103. doi: 10.1016/j.pnpbp.2016.03.010.

American Psychiatric Association (ed.) (2013) *Diagnostic and statistical manual of mental disorders: DSM-5*. 5th ed. Washington, D.C: American Psychiatric Association.

Ashraf, S. I. *et al.* (2006) 'Synaptic Protein Synthesis Associated with Memory Is Regulated by the RISC Pathway in Drosophila', *Cell*, 124(1), pp. 191–205. doi: 10.1016/j.cell.2005.12.017.

Barbato, C. *et al.* (2007) 'Dicer expression and localization in post-mitotic neurons', *Brain Research*, 1175, pp. 17–27. doi: 10.1016/j.brainres.2007.07.088.

Baudry, A. *et al.* (2010) 'MiR-16 Targets the Serotonin Transporter: A New Facet for Adaptive Responses to Antidepressants', *Science*, 329(5998), pp. 1537–1541. doi: 10.1126/science.1193692.

Baxter, A. J. *et al.* (2013) 'Global prevalence of anxiety disorders: a systematic review and meta-regression', *Psychological Medicine*, 43(5), pp. 897–910. doi: 10.1017/S003329171200147X.

Becker, K. G. *et al.* (2004) 'The Genetic Association Database', *Nature Genetics*, 36(5), pp. 431–432. doi: 10.1038/ng0504-431.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing', *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), pp. 289–300.

Betancur, C. (2011) 'Etiological heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting', *Brain Research*. (The Emerging Neuroscience of Autism Spectrum Disorders), 1380, pp. 42–77. doi: 10.1016/j.brainres.2010.11.078.

Boudreau, R. L. *et al.* (2014) 'Transcriptome-wide Discovery of microRNA Binding Sites in Human Brain', *Neuron*, 81(2), pp. 294–305. doi: 10.1016/j.neuron.2013.10.062.

Bousman, C. A. *et al.* (2009) 'Genetic association studies of methamphetamine use disorders: A systematic review and synthesis', *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 150B(8), pp. 1025–1049. doi: 10.1002/ajmg.b.30936.

Bredy, T. W. *et al.* (2011) 'MicroRNA regulation of neural plasticity and memory', *Neurobiology of Learning and Memory*, 96(1), pp. 89–94. doi: 10.1016/j.nlm.2011.04.004.

Buechler, M. B. and Turley, S. J. (2018) 'A short field guide to fibroblast function in immunity', *Seminars in Immunology*. (The mesenchymal and myeloid regulation of immunity), 35, pp. 48–58. doi: 10.1016/j.smim.2017.11.001.

Cadet, J. L., Bisagno, V. and Milroy, C. M. (2014) 'Neuropathology of substance use disorders', *Acta Neuropathologica*, 127(1), pp. 91–107. doi: 10.1007/s00401-013-1221-7.

Campbell, M. C. and Tishkoff, S. A. (2008) 'African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping', *Annual Review of Genomics and Human Genetics*, 9(1), pp. 403–433. doi: 10.1146/annurev.genom.9.081307.164258.

Cantor, R. M., Lange, K. and Sinsheimer, J. S. (2010) 'Prioritizing GWAS Results: A Review of Statistical Methods and Recommendations for Their Application', *American Journal of Human Genetics*, 86(1), pp. 6–22. doi: 10.1016/j.ajhg.2009.11.017.

Cao, L. *et al.* (2016) 'Cocaine-mediated autophagy in astrocytes involves sigma 1 receptor, PI3K, m-TOR, Atg5/7, Beclin-1 and induces Type II programed cell death', *Molecular neurobiology*, 53(7), pp. 4417–4430. doi: 10.1007/s12035-015-9377-x.

Cattane, N. *et al.* (2019) 'Identification of a miRNAs signature associated with exposure to stress early in life and enhanced vulnerability for schizophrenia: New insights for the key role of miR-125b-1-3p in neurodevelopmental processes', *Schizophrenia Research*. (Childhood Trauma and Psychosis), 205, pp. 63–75. doi: 10.1016/j.schres.2018.07.030.

Chang, C. C. *et al.* (2015) 'Second-generation PLINK: rising to the challenge of larger and richer datasets', *GigaScience*, 4(1). doi: 10.1186/s13742-015-0047-8.

Chen, E. *et al.* (2013) *Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3637064/ (Accessed: 18 November 2019).

Chen, L. *et al.* (2017) 'MiR-29b-3p promotes chondrocyte apoptosis and facilitates the occurrence and development of osteoarthritis by targeting PGRN', *Journal of Cellular and Molecular Medicine*, 21(12), pp. 3347–3359. doi: 10.1111/jcmm.13237.

Ciccarone, D. (2011) 'Stimulant Abuse: Pharmacology, Cocaine, Methamphetamine, Treatment, Attempts at Pharmacotherapy', *Primary Care: Clinics in Office Practice*. (Substance Abuse in Office-Based Practice), 38(1), pp. 41–58. doi: 10.1016/j.pop.2010.11.004.

Cortijo, S. *et al.* (2014) 'Mapping the Epigenetic Basis of Complex Traits', *Science*, 343(6175), pp. 1145–1148. doi: 10.1126/science.1248127.

Cristino, A. S. *et al.* (2014) 'Neurodevelopmental and neuropsychiatric disorders represent an interconnected molecular system', *Molecular Psychiatry*, 19(3), pp. 294–301. doi: 10.1038/mp.2013.16.

Dayem Ullah, A. Z. *et al.* (2018) 'SNPnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine', *Nucleic Acids Research*, 46(W1), pp. W109–W113. doi: 10.1093/nar/gky399.

Dias, C. *et al.* (2014) 'β-catenin mediates stress resilience through Dicer1/microRNA regulation', *Nature*, 516(7529), pp. 51–55. doi: 10.1038/nature13976.

Dickson, D. A. *et al.* (2018) 'Reduced levels of miRNAs 449 and 34 in sperm of mice and men exposed to early life stress', *Translational Psychiatry*, 8(1), pp. 1–10. doi: 10.1038/s41398-018-0146-2.

Dreyer, J.-L. (2010) 'New insights into the roles of microRNAs in drug addiction and neuroplasticity', *Genome Medicine*, 2(12), p. 92. doi: 10.1186/gm213.

Eipper-Mains, J. E. *et al.* (2011) 'microRNA-Seq reveals cocaine-regulated expression of striatal microRNAs', *RNA*, 17(8), pp. 1529–1543. doi: 10.1261/rna.2775511.

Eller, R., Janga, S. and Walsh, S. (2019) *Odyssey: a semi-automated pipeline for phasing, imputation, and analysis of genome-wide genetic data*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6599316/ (Accessed: 14 June 2020).

Erra Díaz, F., Dantas, E. and Geffner, J. (2018) 'Unravelling the Interplay between Extracellular Acidosis and Immune Cells', *Mediators of Inflammation*, 2018. doi: 10.1155/2018/1218297.

Esnafoglu, E. and Ayyıldız, S. N. (2017) 'Decreased levels of serum fibroblast growth factor-2 in children with autism spectrum disorder', *Psychiatry Research*, 257, pp. 79–83. doi: 10.1016/j.psychres.2017.07.028.

Farhadian, M. *et al.* (2017) 'Assessment of Executive Functions in Methamphetamine-addicted Individuals: Emphasis on Duration of Addiction and Abstinence', *Basic and Clinical Neuroscience*, 8(2), pp. 147–153. doi: 10.18869/nirp.bcn.8.2.147.

Fiumara, F. *et al.* (2015) 'MicroRNA-22 Gates Long-Term Heterosynaptic Plasticity in Aplysia through Presynaptic Regulation of CPEB and Downstream Targets', *Cell Reports*, 11(12), pp. 1866–1875. doi: 10.1016/j.celrep.2015.05.034.

Friedman, R. C. *et al.* (2009) 'Most mammalian mRNAs are conserved targets of microRNAs', *Genome Research*, 19(1), pp. 92–105. doi: 10.1101/gr.082701.108.

García-Cabrerizo, R. and García-Fuster, M. J. (2019) 'Methamphetamine binge administration dose-dependently enhanced negative affect and voluntary drug consumption in rats following prolonged withdrawal: role of hippocampal FADD', *Addiction Biology*, 24(2), pp. 239–250. doi: 10.1111/adb.12593.

Gay, G. R. (1982) 'Clinical management of acute and chronic cocaine poisoning', *Annals of Emergency Medicine*, 11(10), pp. 562–572. doi: 10.1016/s0196-0644(82)80434-4.

Ge, X. *et al.* (2018) 'Increased miR-21-3p in Injured Brain Microvascular Endothelial Cells after Traumatic Brain Injury Aggravates Blood–Brain Barrier Damage by Promoting Cellular Apoptosis and Inflammation through Targeting MAT2B', *Journal of Neurotrauma*. Mary Ann Liebert, Inc., publishers, 36(8), pp. 1291–1305. doi: 10.1089/neu.2018.5728.

Gelernter, J. *et al.* (2014) 'Genome-wide association study of cocaine dependence and related traits: FAM53B identified as a risk gene', *Molecular Psychiatry*, 19(6), pp. 717–723. doi: 10.1038/mp.2013.99.

González, B. *et al.* (2016) 'Methamphetamine blunts Ca2+ currents and excitatory synaptic transmission through D1/5 receptor-mediated mechanisms in the mouse medial prefrontal cortex', *Addiction Biology*, 21(3), pp. 589–602. doi: 10.1111/adb.12249.

Guha, P., Harraz, M. M. and Snyder, S. H. (2016) 'Cocaine elicits autophagic cytotoxicity via a nitric oxide-GAPDH signaling cascade', *Proceedings of the National Academy of Sciences of the United States of America*, 113(5), pp. 1417–1422. doi: 10.1073/pnas.1524860113.

Haflidadóttir, B. S. *et al.* (2013) 'Upregulation of miR-96 Enhances Cellular Proliferation of Prostate Cancer Cells through FOXO1', *PLOS ONE*. Public Library of Science, 8(8), p. e72400. doi: 10.1371/journal.pone.0072400.

Hajarnis, S., Lakhia, R. and Patel, V. (2020) *Figure 1. [A schematic of miRNA biogenesis...]. - Polycystic Kidney Disease - NCBI Bookshelf*. Available at: https://www.ncbi.nlm.nih.gov/books/NBK373371/figure/fig13_1/#__NBK373371_dtls__ (Accessed: 17 March 2020).

Han, Z. *et al.* (2017) 'MiR-21/PTEN Axis Promotes Skin Wound Healing by Dendritic Cells Enhancement', *Journal of Cellular Biochemistry*, 118(10), pp. 3511–3519. doi: 10.1002/jcb.26026.

Hando, J., Topp, L. and Hall, W. (1997) 'Amphetamine-related harms and treatment preferences of regular amphetamine users in Sydney, Australia', *Drug and Alcohol Dependence*, 46(1–2), pp. 105–113. doi: 10.1016/s0376-8716(97)00051-3.

Hariharan Nirmala *et al.* (2010) 'Deacetylation of FoxO by Sirt1 Plays an Essential Role in Mediating Starvation-Induced Autophagy in Cardiac Myocytes', *Circulation Research*. American Heart Association, 107(12), pp. 1470–1482. doi: 10.1161/CIRCRESAHA.110.227371.

Harms, R. *et al.* (2012) 'Methamphetamine Administration Targets Multiple Immune Subsets and Induces Phenotypic Alterations Suggestive of Immunosuppression', *PLOS ONE*, 7(12), p. e49897. doi: 10.1371/journal.pone.0049897.

Hart, A. B., de Wit, H. and Palmer, A. A. (2012) 'Genetic Factors Modulating the Response to Stimulant Drugs in Humans', *Current topics in behavioral neurosciences*, 12, pp. 537–577. doi: 10.1007/7854_2011_187.

He, W. *et al.* (2017) 'MiR-21 is required for anti-tumor immune response in mice: an implication for its bi-directional roles', *Oncogene*. Nature Publishing Group, 36(29), pp. 4212–4223. doi: 10.1038/onc.2017.62.

He, Y. *et al.* (2015) 'The effects of microRNA on the absorption, distribution, metabolism and excretion of drugs', *British Journal of Pharmacology*, 172(11), pp. 2733–2747. doi: 10.1111/bph.12968.

Hu, C. *et al.* (2020) 'The solute carrier transporters and the brain: Physiological and pharmacological implications', *Asian Journal of Pharmaceutical Sciences*. (Emerging Role of Transporters in Drug Interaction and Delivery), 15(2), pp. 131–144. doi: 10.1016/j.ajps.2019.09.002.

Hu, Z. and Li, Z. (2017) 'miRNAs in synapse development and synaptic plasticity', *Current Opinion in Neurobiology*. (Molecular Neuroscience), 45, pp. 24–31. doi: 10.1016/j.conb.2017.02.014.

Hunt, S. E. *et al.* (2018) 'Ensembl variation resources', *Database*, 2018. doi: 10.1093/database/bay119.

Hyman, S. E., Malenka, R. C. and Nestler, E. J. (2006) 'Neural mechanisms of addiction: the role of reward-related learning and memory', *Annual Review of Neuroscience*, 29, pp. 565–598. doi: 10.1146/annurev.neuro.29.051605.113009.

Ikeda, M. *et al.* (2013) 'Evidence for Shared Genetic Risk Between Methamphetamine-Induced Psychosis and Schizophrenia', *Neuropsychopharmacology*, 38(10), pp. 1864–1870. doi: 10.1038/npp.2013.94.

Issler, O. and Chen, A. (2015) 'Determining the role of microRNAs in psychiatric disorders', *Nature Reviews Neuroscience*, 16(4), pp. 201–212. doi: 10.1038/nrn3879.

Jensen, K. P. (2016) 'A Review of Genome-Wide Association Studies of Stimulant and Opioid Use Disorders', *Molecular Neuropsychiatry*, 2(1), pp. 37–45. doi: 10.1159/000444755.

Jirtle, R. L. and Skinner, M. K. (2007) 'Environmental epigenomics and disease susceptibility', *Nature Reviews Genetics*, 8(4), pp. 253–262. doi: 10.1038/nrg2045.

Karagkouni, D. *et al.* (2018) 'DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA–gene interactions', *Nucleic Acids Research*, 46(D1), pp. D239–D245. doi: 10.1093/nar/gkx1141.

Kays, J. S. and Yamamoto, B. K. (2019) 'Evaluation of Microglia/Macrophage Cells from Rat Striatum and Prefrontal Cortex Reveals Differential Expression of Inflammatory-Related mRNA after Methamphetamine', *Brain Sciences*, 9(12). doi: 10.3390/brainsci9120340.

Kendler, K. S. *et al.* (2005) 'Genetic and environmental influences on illicit drug use and tobacco use across birth cohorts', *Psychological Medicine*, 35(9), pp. 1349–1356. doi: 10.1017/S0033291705004964.

Kendler, K. S., Karkowski, L. and Prescott, C. A. (1999) 'Hallucinogen, opiate, sedative and stimulant use and abuse in a population-based sample of female twins', *Acta Psychiatrica Scandinavica*, 99(5), pp. 368–376. doi: 10.1111/j.1600-0447.1999.tb07243.x.

Khamisipour, G. *et al.* (2018) 'Knockdown of microRNA-29a regulates the expression of apoptosis-related genes in MCF-7 breast carcinoma cells', *Molecular and Clinical Oncology*. Spandidos Publications, 8(2), pp. 362–369. doi: 10.3892/mco.2017.1528.

Kloosterman, W. P. and Plasterk, R. H. A. (2006) 'The Diverse Functions of MicroRNAs in Animal Development and Disease', *Developmental Cell*, 11(4), pp. 441–450. doi: 10.1016/j.devcel.2006.09.009.

Krol, J., Loedige, I. and Filipowicz, W. (2010) 'The widespread regulation of microRNA biogenesis, function and decay', *Nature Reviews. Genetics*, 11(9), pp. 597–610. doi: 10.1038/nrg2843.

Kunkanjanawan, T. *et al.* (2016) 'miR-196a Ameliorates Cytotoxicity and Cellular Phenotype in Transgenic Huntington's Disease Monkey Neural Cells', *PLoS ONE*, 11(9). doi: 10.1371/journal.pone.0162788.

Lahiri, D. K. and Maloney, B. (2012) 'Gene × environment interaction by a longitudinal epigenome-wide association study (LEWAS) overcomes limitations of genome-wide association study (GWAS)', *Epigenomics*, 4(6), pp. 685–699. doi: 10.2217/epi.12.60.

Lanesman, T. H. *et al.* (2019) 'Correlates and predictors of impulsivity in adults with methamphetamine use disorder', *Journal of Substance Use*, 24(4), pp. 361–367. doi: 10.1080/14659891.2019.1572803.

Lee, Y. *et al.* (2003) 'The nuclear RNase III Drosha initiates microRNA processing', *Nature*, 425(6956), pp. 415–419. doi: 10.1038/nature01957.

Li, H. *et al.* (2018) 'Expression of microRNAs in the serum exosomes of methamphetamine-dependent rats vs. ketamine-dependent rats', *Experimental and Therapeutic Medicine*, 15(4), pp. 3369–3375. doi: 10.3892/etm.2018.5814.

Lippman, J. and Dunaevsky, A. (2005) 'Dendritic spine morphogenesis and plasticity', *Journal of Neurobiology*, 64(1), pp. 47–57. doi: 10.1002/neu.20149.

Liu, M. *et al.* (2016) 'MicroRNA-29c overexpression inhibits proliferation and promotes apoptosis and differentiation in P19 embryonal carcinoma cells', *Gene*, 576(1, Part 2), pp. 304–311. doi: 10.1016/j.gene.2015.10.038.

Lopez, J. P. *et al.* (2014) 'miR-1202 is a primate-specific and brain-enriched microRNA involved in major depression and antidepressant treatment', *Nature Medicine*, 20(7), pp. 764–768. doi: 10.1038/nm.3582.

Luikart, B. W., Perederiy, J. V. and Westbrook, G. L. (2012) 'Dentate gyrus neurogenesis, integration and microRNAs', *Behavioural Brain Research*, 227(2), pp. 348–355. doi: 10.1016/j.bbr.2011.03.048.

Luo, S. and Rubinsztein, D. C. (2013) 'Bcl2l11/Bim', *Autophagy*. Taylor & Francis, 9(1), pp. 104–105. doi: 10.4161/auto.22399.

Luoni, A. and Riva, M. A. (2016) 'MicroRNAs and psychiatric disorders: From aetiology to treatment', *Pharmacology & Therapeutics*, 167, pp. 13–27. doi: 10.1016/j.pharmthera.2016.07.006.

Lv, K. *et al.* (2018) 'MiR-22-3p Regulates Cell Proliferation and Inhibits Cell Apoptosis through Targeting the eIF4EBP3 Gene in Human Cervical Squamous Carcinoma Cells', *International Journal of Medical Sciences*, 15(2), pp. 142–152. doi: 10.7150/ijms.21645.

Ma, J. *et al.* (2016) 'microRNA-22 attenuates neuronal cell apoptosis in a cell model of traumatic brain injury', *American Journal of Translational Research*, 8(4), pp. 1895–1902.

Ma, J. and Amos, C. I. (2010) 'Theoretical Formulation of Principal Components Analysis to Detect and Correct for Population Stratification', *PLOS ONE*. Public Library of Science, 5(9), p. e12510. doi: 10.1371/journal.pone.0012510.

Magill, S. T. *et al.* (2010) 'microRNA-132 regulates dendritic growth and arborization of newborn neurons in the adult hippocampus', *Proceedings of the National Academy of Sciences*, 107(47), pp. 20382–20387. doi: 10.1073/pnas.1015691107.

Mahajan, S. D. *et al.* (2006) 'Methamphetamine modulates gene expression patterns in monocyte derived mature dendritic cells: implications for HIV-1 pathogenesis', *Molecular Diagnosis & Therapy*, 10(4), pp. 257–269. doi: 10.1007/BF03256465.

Marchini, J. and Howie, B. (2010) *Genotype Imputation for Genome-Wide Association Studies*, *Nature reviews. Genetics*. Nat Rev Genet. doi: 10.1038/nrg2796.

Martinez, L. R. *et al.* (2009) 'Methamphetamine enhances histoplasmosis by immunosuppression of the host', *The Journal of Infectious Diseases*, 200(1), pp. 131–141. doi: 10.1086/599328.

Martinez, N. J. and Walhout, A. J. M. (2009) 'The interplay between transcription factors and microRNAs in genome-scale regulatory networks', *BioEssays*, 31(4), pp. 435–445. doi: 10.1002/bies.200800212.

Martire, S., Mosca, L. and d'Erme, M. (2015) 'PARP-1 involvement in neurodegeneration: A focus on Alzheimer's and Parkinson's diseases', *Mechanisms of Ageing and Development*, 146–148, pp. 53–64. doi: 10.1016/j.mad.2015.04.001.

McClellan, J. M., Susser, E. and King, M.-C. (2007) 'Schizophrenia: a common disease caused by multiple rare alleles', *The British Journal of Psychiatry: The Journal of Mental Science*, 190, pp. 194–199. doi: 10.1192/bjp.bp.106.025585.

McGregor, N. W. *et al.* (2019) 'Genome-wide analyses indirectly implicate miRNA regulatory mechanisms in Obsessive-compulsive Disorder psychopathology', *bioRxiv*, p. 601401. doi: 10.1101/601401.

McKetin, R. *et al.* (2008) 'Characteristics and harms associated with injecting versus smoking methamphetamine among methamphetamine treatment entrants', *Drug and Alcohol Review*, 27(3), pp. 277–285. doi: 10.1080/09595230801919486.

Mendes-Silva, A. P. *et al.* (2016) 'Shared Biologic Pathways Between Alzheimer Disease and Major Depression: A Systematic Review of MicroRNA Expression Studies', *The American Journal of Geriatric Psychiatry*, 24(10), pp. 903–912. doi: 10.1016/j.jagp.2016.07.017.

Miller, S. A., Dykes, D. D. and Polesky, H. F. (1988) 'A simple salting out procedure for extracting DNA from human nucleated cells.', *Nucleic Acids Research*. Oxford University Press, 16(3), p. 1215. doi: 10.1093/nar/16.3.1215.

Moszczynska, A. (2016) 'Chapter 17 - Methamphetamine Neurotoxicity and the Ubiquitin–Proteasome System', in Preedy, V. R. (ed.) *Neuropathology of Drug Addictions and Substance Misuse*. San Diego: Academic Press, pp. 181–188. doi: 10.1016/B978-0-12-800212-4.00017-0.

Need, A. C. and Goldstein, D. B. (2010) 'Whole genome association studies in complex diseases: where do we stand?', *Dialogues in Clinical Neuroscience*, 12(1), pp. 37–46.

O'Connor, R. M. *et al.* (2016) 'All Roads Lead to the miRNome: miRNAs Have a Central Role in the Molecular Pathophysiology of Psychiatric Disorders', *Trends in Pharmacological Sciences*, 37(12), pp. 1029–1044. doi: 10.1016/j.tips.2016.10.004.

Pandey, A. *et al.* (2015) 'Transactivation of P53 by cypermethrin induced miR-200 and apoptosis in neuronal cells', *Toxicology Research*. Oxford Academic, 4(6), pp. 1578–1586. doi: 10.1039/c5tx00200a.

Papageorgiou, M. *et al.* (2019) 'Methamphetamine and its immune-modulating effects', *Maturitas*, 121, pp. 13–21. doi: 10.1016/j.maturitas.2018.12.003.

Peltzer, K. and Phaswana-Mafuya, N. (2018) 'Drug use among youth and adults in a population-based survey in South Africa', *The South African Journal of Psychiatry : SAJP : the Journal of the Society of Psychiatrists of South Africa*, 24. doi: 10.4102/sajpsychiatry.v24i0.1139.

Popejoy, A. B. and Fullerton, S. M. (2016) 'Genomics is failing on diversity', *Nature*, 538(7624), pp. 161–164. doi: 10.1038/538161a.

Potula, R. *et al.* (2018) 'Methamphetamine alters T cell cycle entry and progression: role in immune dysfunction', *Cell Death Discovery*. Nature Publishing Group, 4(1), pp. 1–13. doi: 10.1038/s41420-018-0045-6.

Price, A. L. *et al.* (2006) 'Principal components analysis corrects for stratification in genome-wide association studies', *Nature Genetics*. Nature Publishing Group, 38(8), pp. 904–909. doi: 10.1038/ng1847.

Prom-Wormley, E. C. *et al.* (2017) 'The genetic epidemiology of substance use disorder: A review', *Drug and Alcohol Dependence*, 180, pp. 241–259. doi: 10.1016/j.drugalcdep.2017.06.040.

Quansah, E. and Karikari, T. K. (2015) 'Motor Neuron Diseases in Sub-Saharan Africa: The Need for More Population-Based Studies', *BioMed Research International*, 2015, pp. 1–9. doi: 10.1155/2015/298409.

Quansah, E. and McGregor, N. W. (2018) 'Towards diversity in genomics: The emergence of neurogenomics in Africa?', *Genomics*, 110(1), pp. 1–9. doi: 10.1016/j.ygeno.2017.07.009.

R Core Team (2019) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/.

Ramsay, M. *et al.* (2011) 'Africa: the next frontier for human disease gene discovery?', *Human Molecular Genetics*, 20(R2), pp. R214–R220. doi: 10.1093/hmg/ddr401.

Reddy-Alla, S. *et al.* (2017) 'Stable Positioning of Unc13 Restricts Synaptic Vesicle Fusion to Defined Release Sites to Promote Synchronous Neurotransmission', *Neuron*, 95(6), pp. 1350-1364.e12. doi: 10.1016/j.neuron.2017.08.016.

Richards, J. R. and Laurin, E. G. (2020) 'Methamphetamine Toxicity', in *StatPearls*. Treasure Island (FL): StatPearls Publishing. Available at: http://www.ncbi.nlm.nih.gov/books/NBK430895/ (Accessed: 12 March 2020).

Romanelli, F. and Smith, K. M. (2006) 'Clinical effects and management of methamphetamine abuse', *Pharmacotherapy*, 26(8), pp. 1148–1156. doi: 10.1592/phco.26.8.1148.

Roshyara, N. *et al.* (2014) *Impact of pre-imputation SNP-filtering on genotype imputation results*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4236550/ (Accessed: 14 June 2020).

Saba, R. and Schratt, G. M. (2010) 'MicroRNAs in neuronal development, function and dysfunction', *Brain Research*, 1338, pp. 3–13. doi: 10.1016/j.brainres.2010.03.107.

Schaefer, A. *et al.* (2007) 'Cerebellar neurodegeneration in the absence of microRNAs', *Journal of Experimental Medicine*, 204(7), pp. 1553–1558. doi: 10.1084/jem.20070823.

Schurz, H. *et al.* (2019) 'Evaluating the Accuracy of Imputation Methods in a Five-Way Admixed Population', *Frontiers in Genetics*, 10. doi: 10.3389/fgene.2019.00034.

Shaukat, S. S., Rao, T. A. and Khan, M. A. (2016) 'Impact of sample size on principal component analysis ordination of an environmental data set: effects on eigenstructure', *Ekológia (Bratislava)*. Sciendo, 35(2), pp. 173–190. doi: 10.1515/eko-2016-0014.

Shi, Y. *et al.* (2017) 'Overexpression of microRNA-96-5p inhibits autophagy and apoptosis and enhances the proliferation, migration and invasiveness of human breast cancer cells', *Oncology Letters*. Spandidos Publications, 13(6), pp. 4402–4412. doi: 10.3892/ol.2017.6025.

Sim, M. S. *et al.* (2017) 'MicroRNA expression signature of methamphetamine use and addiction in the rat nucleus accumbens', *Metabolic Brain Disease*, 32(6), pp. 1767–1783. doi: 10.1007/s11011-017-0061-x.

Sim, S.-E. *et al.* (2016) 'The Brain-Enriched MicroRNA miR-9-3p Regulates Synaptic Plasticity and Memory', *Journal of Neuroscience*. Society for Neuroscience, 36(33), pp. 8641–8652. doi: 10.1523/JNEUROSCI.0630-16.2016.

Simón-Sánchez, J. and Singleton, A. (2008) 'Genome-wide association studies in neurological disorders', *The Lancet. Neurology*, 7(11), pp. 1067–1072. doi: 10.1016/S1474-4422(08)70241-2.

Singer, B. F. *et al.* (2009) 'Amphetamine-induced changes in dendritic morphology in rat forebrain correspond to associative drug conditioning rather than non-associative drug sensitization', *Biological psychiatry*, 65(10), pp. 835–840. doi: 10.1016/j.biopsych.2008.12.020.

Skol, A. D. *et al.* (2007) 'Optimal designs for two-stage genome-wide association studies', *Genetic Epidemiology*, 31(7), pp. 776–788. doi: 10.1002/gepi.20240.

Smyth, L. A. *et al.* (2015) 'MicroRNAs affect dendritic cell function and phenotype', *Immunology*, 144(2), pp. 197–205. doi: 10.1111/imm.12390.

Sommers, I., Baskin, D. and Baskin-Sommers, A. (2006) 'Methamphetamine use among young adults: Health and social consequences', *Addictive Behaviors*. Netherlands: Elsevier Science, 31(8), pp. 1469–1476. doi: 10.1016/j.addbeh.2005.10.004.

Sriram, U. *et al.* (2016) 'Methamphetamine induces trace amine-associated receptor 1 (TAAR1) expression in human T lymphocytes: role in immunomodulation', *Journal of Leukocyte Biology*, 99(1), pp. 213–223. doi: 10.1189/jlb.4A0814-395RR.

Stahl, E. A. *et al.* (2018) 'Genome-wide association study identifies 30 Loci Associated with Bipolar Disorder', *bioRxiv*. Cold Spring Harbor Laboratory, p. 173062. doi: 10.1101/173062.

Stornetta, R. L. and Zhu, J. J. (2011) 'Ras and Rap Signaling in Synaptic Plasticity and Mental Disorders', *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry*, 17(1), pp. 54–78. doi: 10.1177/1073858410365562.

Stranger, B. E., Stahl, E. A. and Raj, T. (2011) 'Progress and Promise of Genome-Wide Association Studies for Human Complex Trait Genetics', *Genetics*, 187(2), pp. 367–383. doi: 10.1534/genetics.110.120907.

Substance Abuse and Mental Health Services Administration, U. (2016) *Substance Use Disorders*. Substance Abuse and Mental Health Services Administration (US). Available at: https://www.ncbi.nlm.nih.gov/books/NBK519702/ (Accessed: 20 February 2020).

Sun, L. *et al.* (2019) 'The role of chaperone-mediated autophagy in neurotoxicity induced by alpha-synuclein after methamphetamine exposure', *Brain and Behavior*, 9(8). doi: 10.1002/brb3.1352.

Sun, Q. *et al.* (2019) 'MicroRNA-190 alleviates neuronal damage and inhibits neuroinflammation via Nlrp3 in MPTP-induced Parkinson's disease mouse model', *Journal of Cellular Physiology*, 234(12), pp. 23379–23387. doi: 10.1002/jcp.28907.

Tan, E.-K. (2007) 'The role of common genetic risk variants in Parkinson disease', *Clinical Genetics*, 72(5), pp. 387–393. doi: 10.1111/j.1399-0004.2007.00890.x.

Tehrani, A. M. *et al.* (2019) 'Methamphetamine induces neurotoxicity-associated pathways and stereological changes in prefrontal cortex', *Neuroscience Letters*, 712, p. 134478. doi: 10.1016/j.neulet.2019.134478.

The Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium (2011) 'Genome-wide association study identifies five new schizophrenia loci', *Nature Genetics*, 43(10), pp. 969–976. doi: 10.1038/ng.940.

Thomas, E. *et al.* (2016) *Methamphetamine-induced psychosis: Clinical features, treatment modalities and outcomes*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6138095/ (Accessed: 18 November 2019).

Treistman, S. N. and Martin, G. E. (2009) 'BK Channels: mediators and models for alcohol tolerance', *Trends in Neurosciences*, 32(12), pp. 629–637. doi: 10.1016/j.tins.2009.08.001.

Turner, S. *et al.* (2011) 'Quality Control Procedures for Genome Wide Association Studies', *Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.]*, CHAPTER, p. Unit1.19. doi: 10.1002/0471142905.hg0119s68.

Turner, S. D. (2014) 'qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots', *bioRxiv*, p. 005165. doi: 10.1101/005165.

Uhl, G. R. *et al.* (2008) 'Genome-Wide Association for Methamphetamine Dependence: Convergent Results From 2 Samples', *Archives of General Psychiatry*, 65(3), pp. 345–355. doi: 10.1001/archpsyc.65.3.345.

Wang, B. *et al.* (2016) 'Essential control of mitochondrial morphology and function by chaperone-mediated autophagy through degradation of PARK7', *Autophagy*. Taylor & Francis, 12(8), pp. 1215–1228. doi: 10.1080/15548627.2016.1179401.

Wang, Q. *et al.* (2017) 'Methamphetamine induces hepatotoxicity via inhibiting cell division, arresting cell cycle and activating apoptosis: In vivo and in vitro studies', *Food and Chemical Toxicology*, 105, pp. 61–72. doi: 10.1016/j.fct.2017.03.030.

Wang, X. *et al.* (2016) 'Discovery and validation of sub-threshold genome-wide association study loci using epigenomic signatures', *eLife*, 5. doi: 10.7554/eLife.10557.

Wang, Y. *et al.* (2012) 'ROS-induced mitochondrial depolarization initiates PARK2/PARKIN-dependent mitochondrial degradation by autophagy', *Autophagy*. Taylor & Francis, 8(10), pp. 1462–1476. doi: 10.4161/auto.21211.

Watt, M. H. *et al.* (2015) 'The mental health experiences and needs of methamphetamine users in Cape Town: A mixed methods study', *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde*, 105(8), pp. 685–688.

Wei, N. *et al.* (2016) 'MicroRNA-9 Mediates the Cell Apoptosis by Targeting Bcl2l11 in Ischemic Stroke', *Molecular Neurobiology*, 53(10), pp. 6809–6817. doi: 10.1007/s12035-015-9605-4.

Wei, R. *et al.* (2015) 'Hydrogen Suppresses Hypoxia/Reoxygenation-Induced Cell Death in Hippocampal Neurons Through Reducing Oxidative Stress', *Cellular Physiology and Biochemistry*. Karger Publishers, 36(2), pp. 585–598. doi: 10.1159/000430122.

Wei, R. *et al.* (2016) 'miR-140-5p attenuates chemotherapeutic drug-induced cell death by regulating autophagy through inositol 1,4,5-trisphosphate kinase 2 (IP3k2) in human osteosarcoma cells', *Bioscience Reports*. Portland Press, 36(5). doi: 10.1042/BSR20160238.

Wen, D. *et al.* (2019) 'Effects of Molecular Hydrogen on Methamphetamine-Induced Neurotoxicity and Spatial Memory Impairment', *Frontiers in Pharmacology*, 10. doi: 10.3389/fphar.2019.00823.

Weybright, E. H. *et al.* (2016) 'The state of methamphetamine ("tik") use among youth in the Western Cape, South Africa', *South African Medical Journal*, 106(11), pp. 1125–1128.

Wightman, B., Ha, I. and Ruvkun, G. (1993) 'Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans', *Cell*, 75(5), pp. 855–862. doi: 10.1016/0092-8674(93)90530-4.

Xue, Q. *et al.* (2016) 'miR-9 and miR-124 synergistically affect regulation of dendritic branching via the AKT/GSK3β pathway by targeting Rap2a', *Scientific Reports*. Nature Publishing Group, 6(1), p. 26781. doi: 10.1038/srep26781.

Yang, J. *et al.* (2011) 'Genomic inflation factors under polygenic inheritance', *European Journal of Human Genetics*, 19(7), pp. 807–812. doi: 10.1038/ejhg.2011.39.

Yi, R. *et al.* (2003) 'Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs', *Genes & Development*, 17(24), pp. 3011–3016. doi: 10.1101/gad.1158803.

Yu, G. *et al.* (2019) 'MiR-142a-3p and miR-155-5p reduce methamphetamine-induced inflammation: Role of the target protein Peli1', *Toxicology and Applied Pharmacology*, 370, pp. 145–153. doi: 10.1016/j.taap.2019.03.019.

Yu, S. *et al.* (2015) *Recent Advances in Methamphetamine Neurotoxicity Mechanisms and Its Molecular Pathophysiology*, *Behavioural Neurology*. doi: 10.1155/2015/103969.

Zhang, C. *et al.* (2019) 'miRNA-mRNA regulatory network analysis of mesenchymal stem cell treatment in cisplatin-induced acute kidney injury identifies roles for miR-210/Serpine1 and miR-378/Fos in regulating inflammation', *Molecular Medicine Reports*. Spandidos Publications, 20(2), pp. 1509–1522. doi: 10.3892/mmr.2019.10383.

Zhang, F. *et al.* (2018) 'miR-let-7e and miR-34a mediate apoptosis induced by ischemic hypoxia of cardiomyocytes in myocardial infarction models', *Int. J. Clin. Exp. Med.*, 11, pp. 9396–9403.

Zhang, K. *et al.* (2016) 'miR-181a is a negative regulator of GRIA2 in methamphetamine-use disorder', *Scientific Reports*, 6. doi: 10.1038/srep35691.

Zhang, K., Jing, X. and Wang, G. (2016) 'MicroRNAs as regulators of drug abuse and immunity', *Central-European Journal of Immunology*, 41(4), pp. 426–434. doi: 10.5114/ceji.2016.65142.

Zhang, Y. *et al.* (2015) 'Dopamine Receptor D2 and Associated microRNAs Are Involved in Stress Susceptibility and Resistance to Escitalopram Treatment', *International Journal of Neuropsychopharmacology*, 18(8). doi: 10.1093/ijnp/pyv025.

Zhang, Y. *et al.* (2016) *Mir143-BBC3 cascade reduces microglial survival via interplay between apoptosis and autophagy: Implications for methamphetamine-mediated neurotoxicity*. Available at: http://europepmc.org/article/PMC/5082785 (Accessed: 20 February 2020).

Zhao, J. *et al.* (2019) 'Different roles of Rac1 in the acquisition and extinction of methamphetamine-associated contextual memory in the nucleus accumbens', *Theranostics*, 9(23), pp. 7051–7071. doi: 10.7150/thno.34655.

Zhao, Y. *et al.* (2016) 'Decreased Expression of Plasma MicroRNA in Patients with Methamphetamine (MA) Use Disorder', *Journal of Neuroimmune Pharmacology*, 11(3), pp. 542–548. doi: 10.1007/s11481-016-9671-z.

Zhou, R. *et al.* (2009) 'Evidence for Selective microRNAs and Their Effectors as Common Long-Term Targets for the Actions of Mood Stabilizers', *Neuropsychopharmacology*, 34(6), pp. 1395–1405. doi: 10.1038/npp.2008.131.

Zhu, L. *et al.* (2015) 'Chronic methamphetamine regulates the expression of MicroRNAs and putative target genes in the nucleus accumbens of mice', *Journal of Neuroscience Research*, 93(10), pp. 1600–1610. doi: 10.1002/jnr.23605.

Chapter 6 Appendices

Supplementary Table 1: SNPs passing nominal significance threshold associated with CUD obtained from summary statistics of the study by Gelernter *et al* (2014).

chr	position	a1	a2	marker	p-value
11	62064715	T	C	rs954237	1,91E-08
6	36479195	G	A	rs7745411	2,27E-07
6	36393410	A	T	rs9470273	3,25E-07
4	139808410	T	C	rs11728231	3,26E-07
6	36509052	T	C	rs6899372	3,68E-07
6	36488476	T	C	rs3798473	3,83E-07
6	36512967	G	A	rs7749358	4,12E-07
6	36441322	G	C	rs6457921	4,92E-07
4	139814341	T	C	rs7684118	4,94E-07
4	139820718	C	T	rs13110551	5,84E-07
6	36402588	C	T	rs56763433	6,43E-07
6	36402308	A	G	rs6941593	6,50E-07
6	36418237	C	T	rs9462185	6,52E-07
4	139816169	G	A	rs62322159	6,73E-07
4	139820534	C	T	rs4443229	6,78E-07
6	36397939	G	A	rs59955083	8,03E-07
4	139812554	G	A	rs35389245	8,13E-07
6	36404525	A	G	rs7769765	8,18E-07
4	139816002	A	G	rs62322158	9,10E-07
6	36444456	T	C	rs9296184	1,23E-06
6	36394087	G	A	rs2180929	1,24E-06
1	201104556	C	A	rs11540083	1,31E-06
4	139822028	C	A	rs11726979	1,57E-06
6	36391890	A	C	rs6912117	1,69E-06
8	41070634	G	T	rs7386810	1,77E-06
6	36397996	A	G	rs9470277	1,79E-06
6	36410223	C	T	rs4711450	1,82E-06
6	36495059	A	G	rs60043926	1,86E-06

4	139800416	A	G	rs9998305	1,88E-06
6	36377820	G	A	rs12202691	1,97E-06
6	36413997	A	G	rs12209346	2,05E-06
10	112897070	C	T	rs7069564	2,12E-06
6	36394169	C	T	rs2145751	2,19E-06
4	139814588	G	A	rs6810673	2,28E-06
6	36510627	A	G	rs12211528	2,39E-06
4	139815232	T	A	rs11939466	2,52E-06
6	36403394	A	T	rs1358981	2,72E-06
1	45750607	A	G	rs4376777	2,77E-06
11	62059846	T	C	rs7935840	2,83E-06
1	45740107	G	T	rs1938296	2,84E-06
6	36372525	A	G	rs6457911	2,86E-06
6	36420638	C	T	rs4711452	2,87E-06
6	36375438	G	A	rs6928287	2,92E-06
6	36407994	G	A	rs9470280	2,96E-06
6	36400461	A	C	rs1535510	2,98E-06
15	94163700	G	A	rs4350528	3,04E-06
6	36373794	C	A	rs6457912	3,17E-06
6	36373451	T	A	rs7773253	3,18E-06
6	36372607	C	T	rs6899437	3,19E-06
6	36371316	C	A	rs4713972	3,22E-06
6	36370394	A	C	rs58153398	3,24E-06
6	36479607	A	G	rs73413258	3,27E-06
6	36367516	G	T	rs12660899	3,34E-06
3	124652260	C	G	rs59399252	3,37E-06
10	112900789	G	C	rs11195436	3,37E-06
1	45738749	T	C	rs3128439	3,38E-06
6	36369260	C	T	rs6923337	3,38E-06
6	17745623	G	A	rs6933410	3,41E-06
4	139801497	G	T	rs10011836	3,43E-06
6	36499571	C	T	rs3798471	3,61E-06
6	36504634	T	C	rs6912982	3,62E-06

6	36505590	C	T	rs6939445	3,72E-06
6	36471934	C	T	rs9462190	3,77E-06
6	17749985	C	T	rs66729438	3,80E-06
1	45737438	T	C	rs3128438	3,81E-06
6	36423085	C	T	rs6899421	3,88E-06
6	36478798	C	A	rs933236	3,88E-06
6	36445773	A	G	rs6926796	3,90E-06
6	111194703	C	G	rs2796749	3,91E-06
1	45743629	T	C	rs3128440	3,94E-06
6	36424458	G	A	rs58093537	3,98E-06
10	2494321	C	A	rs11816248	4,02E-06
10	112892136	C	T	rs12244588	4,12E-06
12	129189355	C	T	rs11836815	4,24E-06
6	36422454	T	C	rs73411054	4,25E-06
6	36500277	A	G	rs7763089	4,30E-06
6	36512862	T	G	rs12209196	4,35E-06
6	36458568	C	T	rs1137628	4,39E-06
6	36509376	C	T	rs6920166	4,39E-06
6	36377709	C	T	rs28371845	4,44E-06
6	36457354	C	T	rs1191	4,45E-06
6	36499468	G	A	rs3798472	4,50E-06
6	36499629	C	T	rs3798470	4,50E-06
6	36505050	T	C	rs6917951	4,51E-06
6	36412213	T	C	rs11757842	4,52E-06
6	36508224	A	G	rs6913613	4,52E-06
6	36508322	G	C	rs6457928	4,52E-06
6	36508398	A	T	rs6457929	4,52E-06
6	36495702	C	T	rs9470304	4,56E-06
6	36496912	C	G	rs2267925	4,57E-06
6	36501137	A	C	rs9462193	4,57E-06
6	36439391	A	G	rs4711453	4,62E-06
6	36468616	T	C	rs9470296	4,63E-06
6	36487113	A	G	rs1983781	4,64E-06

6	36488441	T	C	rs3798474	4,64E-06
6	36492824	A	G	rs6911877	4,64E-06
1	45746828	T	C	rs3121736	4,65E-06
6	36480768	A	G	rs9470302	4,65E-06
6	36483390	G	A	rs7738604	4,65E-06
6	36484393	C	A	rs6457923	4,65E-06
6	36486254	C	T	rs7754363	4,65E-06
6	36462007	G	A	rs4598	4,68E-06
6	36466524	C	G	rs9470293	4,68E-06
6	36469751	G	T	rs9470297	4,68E-06
6	36471639	T	C	rs9470300	4,68E-06
6	36472349	A	G	rs2267932	4,68E-06
6	36475854	T	G	rs11756413	4,68E-06
6	36477260	G	A	rs7772800	4,68E-06
6	36477297	A	G	rs7772524	4,68E-06
6	36478186	A	G	rs12190062	4,68E-06
6	36491427	C	T	rs55829962	4,68E-06
6	17746324	G	A	rs12526810	4,75E-06
6	36494228	G	A	rs67018593	4,78E-06
8	41072759	G	T	rs4736944	4,79E-06
6	36422572	G	A	rs73411055	4,81E-06
6	36451115	A	C	rs9470289	4,85E-06
6	36451777	C	T	rs56086408	4,85E-06
6	36454223	A	T	rs2071811	4,85E-06
4	139801629	G	C	rs7441984	4,89E-06
10	2490751	T	C	rs11816528	4,91E-06
6	36454346	A	G	rs2071810	4,93E-06
10	2493626	A	G	rs34666956	4,99E-06
17	29755996	G	T	rs1124838	5,04E-06
6	17744999	C	T	rs141606122	5,06E-06
6	36508036	G	T	rs6913714	5,10E-06
8	41071868	C	T	rs7832533	5,16E-06
4	139799012	C	T	rs6535748	5,19E-06

6	36437136	C	G	rs1570368	5,19E-06
6	36457484	T	C	rs1061632	5,22E-06
6	36363923	A	C	rs9654588	5,24E-06
11	43171177	C	A	rs11530113	5,36E-06
6	111178628	G	A	rs73534830	5,37E-06
6	36379628	G	A	rs12206445	5,45E-06
6	36472709	A	G	rs2267931	5,47E-06
4	139825168	T	C	rs7670204	5,66E-06
6	36536448	C	T	rs2001101	5,66E-06
6	36382044	A	G	rs4713973	5,77E-06
6	36418370	C	T	rs9462186	5,91E-06
6	17744872	G	A	rs67798133	5,96E-06
6	17733918	T	C	rs9396799	6,02E-06
6	36466668	T	C	rs9470294	6,18E-06
10	2492155	G	A	rs7091333	6,26E-06
16	30164488	A	G	rs11865716	6,39E-06
6	36417186	T	C	rs6917981	6,42E-06
10	2486694	A	G	rs11816810	6,49E-06
6	36481223	G	A	rs73413267	6,51E-06
6	36495980	T	C	rs3819760	6,52E-06
6	36380431	G	A	rs73408889	6,58E-06
10	2492655	G	A	rs7092062	6,69E-06
8	41016640	C	T	rs10958663	6,72E-06
1	232987282	C	T	rs2814549	6,73E-06
6	36570366	T	C	rs7344	6,82E-06
6	36354213	G	A	rs7745687	6,91E-06
6	36354214	C	A	rs7745437	6,92E-06
6	17741328	G	C	rs10949442	6,97E-06
6	36431090	G	C	rs6905347	6,98E-06
6	36431469	C	A	rs6905686	6,98E-06
6	36432643	A	T	rs10947607	6,98E-06
6	36429770	A	G	rs6936942	6,99E-06
6	36363487	G	C	rs67313027	7,07E-06

6	36478089	C	T	rs12110954	7,08E-06
6	36441474	T	C	rs6457922	7,09E-06
4	41256085	T	C	rs11722281	7,15E-06
17	2435714	C	T	rs113883698	7,17E-06
1	45671966	A	G	rs4454584	7,26E-06
6	36457018	C	A	rs2146333	7,26E-06
1	45656917	G	T	rs2105458	7,32E-06
6	36446975	G	C	rs2239808	7,34E-06
6	36486917	A	G	rs2267929	7,36E-06
6	36467956	C	A	rs9470295	7,37E-06
6	36492269	C	G	rs3752479	7,37E-06
1	45672783	T	C	rs1938300	7,40E-06
1	45675120	A	G	rs1938301	7,40E-06
1	45662807	A	G	rs2356302	7,41E-06
6	36496635	G	C	rs7742459	7,43E-06
1	45665939	G	A	rs3009974	7,44E-06
1	45679640	A	G	rs1938304	7,57E-06
1	45696149	C	G	rs3009979	7,59E-06
4	41254954	A	C	rs4861386	7,64E-06
6	36410190	G	A	rs6928087	7,94E-06
6	36432892	A	G	rs10947608	7,96E-06
1	45645069	C	T	rs3009969	8,16E-06
6	36361409	G	A	rs9470267	8,21E-06
6	111210654	T	C	rs1963644	8,23E-06
6	36477926	A	G	rs10947609	8,24E-06
6	36453564	T	C	rs57389119	8,27E-06
6	36477679	G	A	rs11755264	8,35E-06
1	45641454	A	C	rs2997465	8,43E-06
6	36506487	A	G	rs6902076	8,49E-06
6	36429088	T	A	rs7752714	8,58E-06
6	36492375	T	G	rs3752478	8,61E-06
10	54530067	A	G	rs4935047	8,73E-06
11	26280766	T	A	rs12270830	8,78E-06

10	54529705	C	T	rs4935046	8,82E-06
6	36451639	G	A	rs9470291	8,85E-06
1	94317477	T	C	rs12023749	8,86E-06
10	54529480	G	A	rs10824793	8,86E-06
6	36422789	A	G	rs6926522	8,89E-06
1	45694796	C	T	rs2997460	8,96E-06
1	45711194	A	C	rs2997458	9,05E-06
1	45722622	A	G	rs1938407	9,13E-06
11	62060451	A	G	rs11600383	9,13E-06
6	36494797	T	C	rs57423505	9,14E-06
1	45639768	A	G	rs3009968	9,28E-06
6	36385744	C	G	rs12181938	9,36E-06
6	36388565	C	T	rs7767377	9,40E-06
6	36464958	G	A	rs4140593	9,40E-06
6	36362441	A	C	rs12663882	9,41E-06
8	76287012	T	C	rs7826193	9,52E-06
6	36499234	A	G	rs3778019	9,61E-06
6	36499860	A	G	rs3822970	9,62E-06
6	36503085	C	G	rs10947611	9,63E-06
6	36505630	A	C	rs2894407	9,64E-06
6	36448766	G	T	rs6902473	9,65E-06
6	36480437	C	T	rs1985423	9,72E-06
8	41070759	T	G	rs57766780	9,74E-06
6	36474141	G	A	rs2300064	9,76E-06
6	36475200	T	C	rs16888690	9,76E-06
6	36496909	C	T	rs2267926	9,76E-06
1	45694672	A	G	rs2997461	9,84E-06
6	36428372	G	A	rs6930496	9,89E-06
1	45733527	T	A	rs2186219	9,96E-06
1	45602076	T	C	rs346687	9,97E-06
6	36429373	A	G	rs7771028	1,00E-05
6	36426183	T	C	rs9767778	1,01E-05
6	111204367	C	G	rs1279593	1,01E-05

8	125835989	T	C	rs10956215	1,02E-05
20	59878158	A	G	rs2427060	1,02E-05
1	45705709	C	A	rs3009981	1,03E-05
6	36449676	C	T	rs3756912	1,04E-05
6	36480499	C	T	rs1985424	1,04E-05
1	45632989	C	T	rs1226913	1,06E-05
4	139825516	T	C	rs7675444	1,06E-05
6	36447287	A	G	rs2239807	1,06E-05
6	36450514	A	G	rs6913693	1,06E-05
6	36451405	T	C	rs6899631	1,06E-05
5	16357489	A	G	rs4701665	1,07E-05
6	36493330	C	T	rs6927896	1,07E-05
6	36495033	A	G	rs57646047	1,07E-05
6	111210420	C	T	rs1027914	1,08E-05
1	45763650	A	C	rs4233496	1,09E-05
6	36482663	T	C	rs2007741	1,09E-05
6	36490479	G	A	rs2267927	1,09E-05
1	45762934	T	G	rs6429561	1,10E-05
6	36415903	G	A	rs10484615	1,10E-05
6	36459466	C	T	rs2187827	1,10E-05
1	45596447	C	G	rs346696	1,11E-05
6	17737690	T	A	rs12199879	1,14E-05
6	36430588	A	G	rs6457920	1,15E-05
12	129188445	T	G	rs11059839	1,15E-05
6	36490133	G	A	rs2267928	1,16E-05
10	54530636	A	G	rs1982266	1,16E-05
16	30336030	A	T	rs7195216	1,17E-05
1	45763753	T	C	rs4233497	1,18E-05
6	36412658	C	T	rs16888552	1,18E-05
16	30334923	G	A	rs7186541	1,18E-05
6	36426271	G	T	rs9767243	1,19E-05
6	36449823	T	A	rs3756911	1,20E-05
1	45607309	C	A	rs1144868	1,22E-05

16	69063947	G	C	rs251101	1,22E-05
6	36468672	C	T	rs59104483	1,24E-05
6	111225883	T	A	rs1279469	1,25E-05
6	36437558	A	T	rs1003249	1,27E-05
1	45591389	T	C	rs415536	1,28E-05
1	45585207	C	T	rs346733	1,30E-05
1	45564873	G	A	rs346731	1,32E-05
1	45577932	T	A	rs346736	1,32E-05
1	45576994	T	G	rs346737	1,34E-05
8	125837132	G	A	rs7825377	1,34E-05
1	45720886	G	T	rs3121732	1,35E-05
6	36361682	G	A	rs2057351	1,35E-05
8	41078439	C	T	rs4736947	1,35E-05
6	17741571	G	A	rs10949443	1,37E-05
8	41071312	T	C	rs10958668	1,37E-05
4	139810185	C	T	rs6815428	1,41E-05
8	76275295	C	G	rs117952276	1,41E-05
11	274397	G	A	rs11246044	1,41E-05
1	45632046	T	C	rs1226914	1,43E-05
6	111283166	C	T	rs9372275	1,43E-05
12	129187152	C	T	rs7968966	1,43E-05
6	111222552	C	T	rs11153258	1,46E-05
8	76286697	C	A	rs58424670	1,46E-05
16	69043117	T	C	rs251104	1,46E-05
6	19613120	A	G	rs1936907	1,47E-05
12	129182178	G	A	rs11836697	1,50E-05
6	36579252	C	T	rs7743396	1,53E-05
6	36597695	A	G	rs11756557	1,53E-05
11	274188	A	G	rs11246043	1,55E-05
16	69047382	T	C	rs251106	1,55E-05
1	45533160	C	T	rs6429557	1,58E-05
6	36498080	C	G	rs67640018	1,58E-05
1	45721934	A	G	rs3121769	1,59E-05

6	36435004	G	C	rs11759996	1,60E-05
1	45620838	G	A	rs346723	1,62E-05
6	36598209	G	A	rs6913530	1,66E-05
19	45938020	T	A	rs9749017	1,68E-05
1	45530509	T	C	rs11211074	1,73E-05
1	45539034	G	A	rs346701	1,73E-05
3	65575607	T	C	rs80336351	1,75E-05
8	76274299	A	T	rs56683582	1,76E-05
8	125830838	G	T	rs7821433	1,76E-05
6	36586070	G	A	rs12208086	1,77E-05
10	132913582	T	C	rs80139688	1,78E-05
9	38720627	C	T	rs12682690	1,83E-05
6	36363556	T	G	rs73408850	1,86E-05
1	45555836	G	A	rs238249	1,88E-05
6	36379057	G	A	rs7746792	1,91E-05
6	36566575	C	T	rs9368943	1,96E-05
15	54527821	C	T	rs12442104	1,96E-05
6	19626762	A	G	rs6900462	1,97E-05
6	36514721	T	G	rs6941918	1,97E-05
1	45523893	T	C	rs3850855	1,98E-05
2	119919322	G	A	rs35992749	1,98E-05
6	111193422	T	C	rs1592206	1,98E-05
6	36368169	G	A	rs6917489	2,01E-05
12	129191924	A	G	rs12370026	2,01E-05
16	30155031	T	C	rs7197001	2,01E-05
16	30154740	A	G	rs9924308	2,02E-05
6	36378487	A	G	rs12190911	2,06E-05
6	36396908	T	G	rs77953031	2,08E-05
3	171061754	G	A	rs73879536	2,09E-05
16	69063842	C	G	rs251102	2,09E-05
7	146406379	T	C	rs2533096	2,10E-05
2	125508650	C	T	rs17393515	2,11E-05
6	111177160	G	A	rs9384757	2,14E-05

8	120246240	G	A	rs7004820	2,15E-05
8	41071463	T	G	rs7815492	2,18E-05
5	13245491	C	T	rs13170612	2,20E-05
4	139814402	A	T	rs6836659	2,21E-05
6	111174173	G	A	rs812439	2,23E-05
4	138612602	G	A	rs1358094	2,24E-05
6	19614926	A	G	rs72836028	2,24E-05
6	19614940	T	C	rs67321156	2,24E-05
10	2480445	A	G	rs12780201	2,24E-05
4	138612780	C	A	rs1818954	2,28E-05
5	16356980	G	A	rs4702141	2,29E-05
10	2521072	C	T	rs35366259	2,29E-05
10	2467597	A	T	rs11812328	2,30E-05
17	10163579	C	A	rs4791397	2,31E-05
4	138613354	A	G	rs16998355	2,34E-05
1	195210587	G	T	rs78003589	2,35E-05
4	138613596	T	G	rs7698043	2,36E-05
11	43158024	G	A	rs11037260	2,36E-05
19	689836	T	G	rs10408751	2,36E-05
4	138613791	G	A	rs7665269	2,38E-05
6	19616157	A	G	rs1936909	2,38E-05
20	43348735	C	A	rs2296530	2,39E-05
17	10163167	A	T	rs4791396	2,44E-05
17	10162744	A	T	rs4791395	2,45E-05
10	2463990	C	T	rs12785032	2,46E-05
12	129193769	T	C	rs11835679	2,47E-05
1	45580884	T	C	rs1226836	2,50E-05
11	43203225	C	T	rs56983842	2,50E-05
4	89724638	T	G	rs17747153	2,54E-05
16	69067538	T	C	rs32853	2,54E-05
20	9977095	T	C	rs11907417	2,54E-05
14	94567767	C	T	rs11852002	2,58E-05
17	10162911	G	A	rs6503303	2,59E-05

16	69062474	T	C	rs251103	2,62E-05
4	138609753	T	C	rs57623344	2,64E-05
4	138610020	C	T	rs72990875	2,64E-05
4	138610358	T	C	rs17049793	2,64E-05
8	76286530	A	C	rs59134671	2,64E-05
16	69078952	C	T	rs12929269	2,65E-05
5	16390793	A	G	rs335954	2,66E-05
17	10162050	C	T	rs12938237	2,66E-05
6	17732028	C	A	rs12528444	2,67E-05
11	34967328	A	G	rs11032943	2,67E-05
4	138616276	A	C	rs56366116	2,68E-05
4	138616339	G	A	rs11946875	2,69E-05
11	34897474	A	C	rs2163377	2,69E-05
11	43157549	A	T	rs11037259	2,71E-05
3	13621556	G	C	rs11925439	2,72E-05
12	129190913	G	A	rs7980709	2,74E-05
16	69061010	G	A	rs173685	2,76E-05
4	89724384	C	A	rs57956471	2,84E-05
10	2479636	T	C	rs55661812	2,84E-05
6	36573822	G	C	rs10947614	2,85E-05
8	125837679	T	G	rs7844066	2,85E-05
1	45515883	C	T	rs10789454	2,86E-05
6	36405555	T	A	rs59440875	2,86E-05
6	153262157	T	C	rs11969689	2,86E-05
5	16290259	A	T	rs7719785	2,87E-05
6	153263149	A	T	rs12110522	2,87E-05
8	41054165	T	A	rs72641826	2,88E-05
12	115846227	G	A	rs755504	2,89E-05
3	102588429	T	G	rs1144780	2,93E-05
5	16302942	C	T	rs3111618	2,93E-05
7	146402912	A	C	rs344448	2,93E-05
10	2455362	C	T	rs12774747	2,98E-05
7	146404783	T	G	rs6945320	2,99E-05

8	41052622	A	G	rs4736799	2,99E-05
8	41064077	C	T	rs4736943	3,05E-05
6	17736154	G	T	rs144501842	3,06E-05
8	41053548	C	T	rs56409352	3,08E-05
6	111195730	G	A	rs781193	3,17E-05
16	30169393	C	G	rs12927190	3,18E-05
8	125838091	C	G	rs7829946	3,21E-05
1	195166147	G	A	rs3904174	3,22E-05
6	36549858	G	A	rs9357218	3,22E-05
11	274780	T	C	rs11246045	3,22E-05
4	139826720	A	G	rs4639023	3,24E-05
6	36353169	G	A	rs6901513	3,24E-05
16	69068340	C	A	rs32854	3,25E-05
6	111231055	T	C	rs1066751	3,27E-05
8	76287412	C	T	rs7842940	3,33E-05
7	81823960	C	T	rs2040922	3,36E-05
16	30326571	G	C	rs8050176	3,43E-05
19	689977	C	T	rs10418290	3,43E-05
11	20611450	A	G	rs4923219	3,45E-05
6	111190694	A	G	rs7749918	3,51E-05
5	13213877	A	G	rs1445730	3,54E-05
7	140061304	C	T	rs71552606	3,59E-05
6	111190258	T	G	rs9689864	3,63E-05
1	195191127	A	T	rs4341350	3,69E-05
2	26398335	A	C	rs10175960	3,72E-05
6	19616733	G	A	rs55779808	3,72E-05
8	41080124	T	G	rs4736950	3,73E-05
6	19617525	C	T	rs7753244	3,79E-05
17	10163569	A	G	rs4791959	3,80E-05
11	43264801	C	T	rs74969066	3,81E-05
20	9970855	G	A	rs148020671	3,83E-05
6	19617378	T	G	rs7773374	3,84E-05
2	105357297	C	G	rs73944963	3,88E-05

3	65558540	T	C	rs17073130	3,93E-05
14	90892568	A	G	rs10467854	3,94E-05
3	65557937	T	C	rs77356972	3,98E-05
6	36551370	A	C	rs9394361	3,99E-05
6	19617303	G	A	rs7753163	4,02E-05
6	36588140	C	T	rs12211825	4,02E-05
6	36526313	G	A	rs12208373	4,09E-05
7	146402347	G	A	rs344451	4,10E-05
6	36554240	A	C	rs72852338	4,11E-05
6	19621821	T	C	rs9358252	4,13E-05
9	38722442	G	A	rs7045549	4,13E-05
8	76285519	G	T	rs76129372	4,14E-05
9	38722510	C	T	rs10973999	4,15E-05
7	5590478	A	C	rs1474444	4,16E-05
9	138503416	C	T	rs872005	4,19E-05
3	13621911	A	G	rs3773288	4,20E-05
4	160324529	A	G	rs2699015	4,20E-05
7	146406691	A	G	rs2533095	4,21E-05
9	38723314	C	T	rs11789223	4,23E-05
19	15588373	G	C	rs6512036	4,23E-05
6	10467382	T	C	rs471034	4,25E-05
3	65537584	C	G	rs17432146	4,27E-05
8	41061145	T	C	rs55737070	4,32E-05
16	69055675	C	A	rs251100	4,32E-05
12	82503402	A	G	rs7952825	4,33E-05
3	13621922	C	G	rs3773287	4,36E-05
6	19620480	T	C	rs2876558	4,37E-05
16	30158757	T	G	rs7199462	4,38E-05
11	43162217	C	G	rs11037264	4,43E-05
10	118048572	T	C	rs2429477	4,46E-05
1	45764336	A	G	rs6429563	4,47E-05
6	19620452	A	G	rs4610552	4,47E-05
6	19620886	C	T	rs7454652	4,47E-05

9	38721044	C	T	rs10814746	4,50E-05
19	15589701	G	T	rs1961562	4,51E-05
1	45764258	C	T	rs6429562	4,52E-05
6	19619957	G	C	rs9356701	4,53E-05
2	119903969	G	C	rs7597287	4,56E-05
11	43160693	C	A	rs75111899	4,56E-05
21	46770227	G	A	rs113813349	4,58E-05
5	16346572	C	T	rs12656374	4,62E-05
6	17744282	A	G	rs10456805	4,63E-05
9	38722798	G	C	rs10974000	4,64E-05
12	77107695	T	C	rs977192	4,64E-05
6	19617841	A	G	rs1120462	4,65E-05
1	160587121	A	C	rs164275	4,71E-05
17	30875670	A	G	rs225200	4,75E-05
6	111221806	A	T	rs1767978	4,78E-05
5	124529455	G	C	rs6895323	4,81E-05
6	19611284	G	T	rs9350167	4,81E-05
2	119907848	A	C	rs17190912	4,83E-05
18	22123735	A	C	rs949316	4,83E-05
6	19611426	T	C	rs9368098	4,86E-05
21	25624249	T	A	rs9985038	4,91E-05
16	80527580	A	G	rs4305034	4,96E-05
11	43157150	G	A	rs11037257	4,99E-05
16	30159107	G	A	rs12919578	5,02E-05
11	43156918	T	C	rs12225970	5,09E-05
1	195158965	A	T	rs4111374	5,10E-05
6	36596784	T	C	rs12197590	5,13E-05
6	148211075	C	T	rs500698	5,15E-05
19	15589855	C	A	rs7245400	5,21E-05
6	111220559	C	G	rs1628454	5,22E-05
12	115846911	A	G	rs2384636	5,23E-05
6	19608592	A	T	rs11965154	5,24E-05
16	30163157	A	G	rs8050401	5,25E-05

16	9598191	T	C	rs10492846	5,30E-05
9	38724957	T	C	rs73454434	5,34E-05
8	131436649	G	A	rs112765450	5,35E-05
10	33006803	A	G	rs145737300	5,41E-05
1	232997556	C	T	rs2795433	5,42E-05
18	65590367	A	T	rs77156166	5,44E-05
21	46766234	C	T	rs113183596	5,46E-05
6	19613709	C	T	rs9348387	5,47E-05
6	19620361	G	C	rs2328484	5,47E-05
7	146410468	A	C	rs344478	5,47E-05
17	30876182	C	T	rs225195	5,47E-05
1	195171687	G	A	rs78626875	5,48E-05
6	111216211	A	T	rs781194	5,48E-05
1	195176614	G	T	rs12042194	5,49E-05
11	43157391	G	A	rs12224770	5,49E-05
4	160330868	T	C	rs2463987	5,51E-05
6	19624278	T	C	rs113808940	5,51E-05
10	2511234	C	T	rs17261406	5,54E-05
3	29397049	G	A	rs57830027	5,57E-05
6	23946004	A	G	rs477570	5,57E-05
6	19613078	T	C	rs1936905	5,59E-05
6	111200228	G	A	rs1279599	5,59E-05
10	2512010	T	C	rs34424047	5,60E-05
6	19612034	A	G	rs9350168	5,62E-05
12	19469732	T	C	rs7961028	5,63E-05
6	19611739	G	A	rs9358248	5,64E-05
9	85806577	T	C	rs10867947	5,64E-05
6	19611231	C	T	rs9350166	5,65E-05
6	19611375	A	G	rs9295444	5,65E-05
6	19619060	A	G	rs11969049	5,67E-05
10	2513746	C	T	rs17176386	5,67E-05
6	19611913	C	A	rs9358249	5,68E-05
6	36598837	G	C	rs9394365	5,69E-05

5	124342918	C	T	rs4472259	5,70E-05
6	19618388	G	C	rs1954541	5,74E-05
5	84088725	G	A	rs73772672	5,76E-05
8	107795350	G	A	rs80209653	5,76E-05
8	130390400	C	T	rs16904057	5,76E-05
5	13189504	T	G	rs79711428	5,78E-05
6	19621304	C	G	rs7774619	5,78E-05
6	111201474	G	A	rs1279597	5,78E-05
6	111213217	A	G	rs811921	5,79E-05
7	146402727	A	G	rs344449	5,80E-05
10	2514506	A	G	rs17261532	5,80E-05
4	89727562	A	G	rs17747350	5,82E-05
4	89727597	C	T	rs73842241	5,82E-05
4	43426370	G	T	rs484707	5,83E-05
5	13191362	T	C	rs1445726	5,84E-05
6	17744733	T	C	rs2093823	5,84E-05
5	64242490	T	C	rs12519598	5,86E-05
17	30876440	C	T	rs225194	5,87E-05
6	111212283	A	G	rs7768897	5,89E-05
16	80522626	A	C	rs4608348	5,91E-05
8	76286553	T	C	rs57864512	5,92E-05
11	43258061	T	C	rs79120789	5,94E-05
1	64814465	C	T	rs1514749	5,96E-05
10	32798947	G	A	rs10740855	5,97E-05
13	109330018	C	T	rs75020132	6,01E-05
1	45571522	A	G	rs148468469	6,02E-05
3	65577277	T	C	rs79996438	6,03E-05
5	124332738	G	A	rs4240386	6,03E-05
5	16329478	C	T	rs80037969	6,04E-05
8	41045576	T	C	rs6474317	6,04E-05
1	191431169	G	A	rs10920948	6,07E-05
6	111203419	A	G	rs1279596	6,09E-05
6	111205405	G	C	rs1279589	6,11E-05

1	45699584	A	G	rs3009980	6,12E-05
6	111194468	A	G	rs2665359	6,12E-05
7	6131169	C	T	rs539130	6,13E-05
9	135389030	A	G	rs12005531	6,13E-05
10	32787808	T	C	rs3006713	6,17E-05
1	191432168	G	A	rs12691510	6,18E-05
6	111208555	G	A	rs2070969	6,19E-05
1	195249444	A	G	rs74517089	6,21E-05
10	2483585	C	T	rs7900071	6,21E-05
10	2521726	C	T	rs34381879	6,21E-05
6	111207944	T	C	rs7757511	6,23E-05
6	19610320	G	A	rs6905926	6,24E-05
6	21922708	T	G	rs12213632	6,24E-05
2	98624110	C	G	rs7586210	6,25E-05
4	160325490	G	A	rs2699017	6,25E-05
6	19617885	A	G	rs1120463	6,35E-05
9	38725996	C	T	rs7028380	6,37E-05
5	124332814	G	A	rs4835856	6,39E-05
18	34182084	T	C	rs713579	6,43E-05
5	13244265	C	T	rs4277916	6,45E-05
6	21923640	G	C	rs9460688	6,47E-05
11	43198588	G	C	rs11037313	6,47E-05
6	21922996	C	G	rs9466214	6,48E-05
6	21923970	C	A	rs9466216	6,48E-05
16	10270314	T	C	rs7499321	6,50E-05
18	34182266	G	A	rs12605734	6,56E-05
1	58004484	C	T	rs1357782	6,63E-05
3	13618204	G	A	rs9874733	6,67E-05
13	77378713	C	T	rs6562967	6,73E-05
4	139809609	C	G	rs13120367	6,74E-05
6	19609834	T	C	rs9766473	6,74E-05
6	111183369	A	G	rs1279601	6,77E-05
12	77110778	A	G	rs1872822	6,80E-05

5	13201390	A	G	rs66715406	6,82E-05
6	19620779	G	A	rs61578935	6,84E-05
1	191428753	G	A	rs6666273	6,95E-05
6	21914756	G	A	rs9460687	7,06E-05
11	69068692	C	T	rs11228614	7,06E-05
20	42810048	A	C	rs13041775	7,07E-05
1	58000711	T	C	rs2793625	7,08E-05
8	39661189	G	A	rs62511202	7,10E-05
6	148209316	T	C	rs552363	7,11E-05
5	16288090	A	T	rs4515299	7,12E-05
10	2463661	G	T	rs7096935	7,13E-05
5	13218713	T	G	rs59938005	7,14E-05
5	7269098	T	C	rs144955043	7,15E-05
10	32925789	G	C	rs71493136	7,17E-05
11	12147829	G	A	rs2279565	7,19E-05
16	55178799	T	C	rs4784498	7,19E-05
5	154893283	G	A	rs11750027	7,21E-05
7	155104401	G	A	rs56085544	7,23E-05
5	89915573	C	T	rs146198134	7,24E-05
5	154900898	T	C	rs11742116	7,25E-05
11	34889989	C	T	rs3962032	7,25E-05
11	34889842	C	T	rs2102475	7,26E-05
8	41078912	T	C	rs4736948	7,28E-05
10	79504515	A	T	rs2028284	7,30E-05
6	36599596	T	C	rs10947615	7,32E-05
19	15591333	A	G	rs3813136	7,32E-05
6	19621732	T	C	rs9358251	7,33E-05
1	58005178	A	C	rs1524716	7,36E-05
3	30662554	A	G	rs9844092	7,39E-05
3	72187591	G	C	rs13320346	7,39E-05
16	10268412	A	G	rs1104068	7,40E-05
2	186325812	A	C	rs79061010	7,44E-05
1	64812821	A	G	rs10493358	7,46E-05

16	10268731	T	C	rs1102972	7,46E-05
10	118048462	T	C	rs2429478	7,49E-05
11	43214380	G	C	rs11037329	7,52E-05
1	195196489	C	T	rs12030743	7,53E-05
4	190064939	C	A	rs13134876	7,55E-05
2	1196202	A	G	rs4266053	7,56E-05
2	186263008	T	A	rs1559529	7,56E-05
1	195252784	A	T	rs12044765	7,57E-05
16	10268330	T	C	rs1650420	7,57E-05
4	160316508	C	T	rs111295346	7,58E-05
1	195253952	C	T	rs2225285	7,60E-05
4	190064912	G	T	rs13135232	7,61E-05
16	80525710	T	G	rs12448525	7,61E-05
2	1178422	A	G	rs28594422	7,63E-05
7	115283613	T	C	rs10251693	7,63E-05
6	36562021	G	C	rs1076931	7,64E-05
17	9671517	A	T	rs2009608	7,70E-05
5	13198707	A	G	rs75517472	7,73E-05
8	136944138	C	A	rs2317536	7,75E-05
16	30172627	A	G	rs13331817	7,75E-05
16	10268446	C	A	rs8055598	7,77E-05
11	43202447	G	A	rs11037318	7,79E-05
13	95446803	G	A	rs792306	7,79E-05
14	22015018	A	C	rs2319863	7,83E-05
19	42155818	A	C	rs11672853	7,83E-05
1	195259453	C	T	rs12028770	7,84E-05
3	11582619	C	T	rs111973213	7,84E-05
6	21920303	A	T	rs11964329	7,85E-05
19	39318875	G	A	rs28496951	7,85E-05
8	41023480	T	G	rs12545637	7,87E-05
14	87346321	A	G	rs7145996	7,88E-05
16	69036125	A	C	rs1593137	7,91E-05
15	94160966	T	A	rs12908655	7,92E-05

4	190065099	G	A	rs13135632	7,94E-05
6	36398229	T	C	rs629721	7,95E-05
6	19613858	C	A	rs9366302	7,96E-05
10	32804322	G	A	rs7075553	7,97E-05
1	105235777	G	A	rs6679989	8,00E-05
18	23967372	C	G	rs7236786	8,02E-05
9	135389261	C	T	rs7028726	8,03E-05
5	124222894	G	C	rs17514306	8,04E-05
5	154831770	G	A	rs2547011	8,04E-05
6	17730382	T	C	rs7757779	8,08E-05
16	89116980	T	C	rs111390307	8,09E-05
19	42166864	G	A	rs9797683	8,10E-05
19	42167191	C	T	rs12462129	8,11E-05
16	10268243	T	C	rs1104474	8,12E-05
10	32832214	G	A	rs2990986	8,17E-05
10	32834576	G	A	rs1831771	8,17E-05
18	3161025	A	G	rs36115329	8,22E-05
6	17729829	C	G	rs10949440	8,23E-05
2	1184107	G	A	rs4971408	8,24E-05
11	270514	C	T	rs17156153	8,24E-05
19	42162709	G	A	rs35633112	8,26E-05
11	43205710	G	A	rs78830907	8,27E-05
11	43205764	A	G	rs77555029	8,27E-05
11	43251862	T	C	rs76228664	8,27E-05
6	111183002	G	T	rs7774455	8,28E-05
11	43205276	C	T	rs56110731	8,28E-05
9	135388415	T	A	rs6597576	8,29E-05
18	22120936	T	C	rs17798101	8,29E-05
10	32827578	C	T	rs7910275	8,30E-05
11	43198223	T	A	rs11037312	8,33E-05
10	107017094	C	T	rs72819929	8,34E-05
11	43251244	C	A	rs79208374	8,38E-05
11	43198684	G	A	rs11037314	8,39E-05

11	43203563	G	A	rs11037320	8,40E-05
10	33001361	T	G	rs7900825	8,43E-05
11	43161000	C	T	rs11037262	8,44E-05
19	42154137	G	A	rs4347734	8,45E-05
6	36553750	C	T	rs12202732	8,47E-05
19	42155152	G	T	rs11671944	8,47E-05
11	275752	G	C	rs12419714	8,50E-05
13	53814652	G	A	rs79820428	8,50E-05
1	233004712	C	T	rs1008536	8,51E-05
6	17729816	T	C	rs10949439	8,51E-05
6	40912035	G	A	rs62397190	8,51E-05
11	43199446	G	A	rs11037315	8,51E-05
19	42157068	C	T	rs917552	8,52E-05
12	77107595	C	A	rs2219734	8,54E-05
6	19619924	T	C	rs9356700	8,57E-05
10	107027306	A	G	rs10884127	8,58E-05
11	43203469	T	C	rs12224893	8,60E-05
11	43211475	A	G	rs77089670	8,61E-05
12	77111991	A	T	rs7315813	8,63E-05
1	161234623	A	C	rs11265577	8,64E-05
2	12964510	C	T	rs17465833	8,69E-05
11	43239926	G	A	rs11037361	8,69E-05
5	16291138	A	G	rs6885358	8,70E-05
11	43211626	C	G	rs77983318	8,70E-05
3	65572675	C	T	rs76273363	8,73E-05
17	77925393	G	A	rs35286723	8,73E-05
3	52516293	C	G	rs6800707	8,75E-05
10	32955086	G	C	rs7899433	8,75E-05
20	15312550	T	A	rs6034159	8,75E-05
1	103322312	C	T	rs11164621	8,78E-05
6	19619036	G	A	rs55679921	8,79E-05
13	22538559	G	A	rs6490688	8,79E-05
11	43209612	T	A	rs57870742	8,80E-05

2	186303530	A	G	rs767037	8,81E-05
6	5822122	T	G	rs1830687	8,82E-05
6	19619700	G	A	rs9368101	8,83E-05
8	41028520	C	T	rs6983729	8,83E-05
10	33080104	T	C	rs7914502	8,85E-05
5	13229090	A	G	rs35637368	8,90E-05
11	43248498	G	A	rs75024144	8,91E-05
19	35741456	T	C	rs2073900	8,91E-05
6	5822043	A	C	rs4370407	8,94E-05
6	5822049	C	T	rs1830688	8,94E-05
6	19625748	A	G	rs10434828	8,95E-05
19	42156168	T	C	rs11880469	8,98E-05
6	21912672	C	T	rs6456458	8,99E-05
4	139808394	A	G	rs11727347	9,01E-05
4	22252771	T	G	rs55738732	9,04E-05
10	32900543	A	G	rs10740858	9,04E-05
8	41036557	T	G	rs11774048	9,05E-05
12	77107338	A	C	rs7953372	9,05E-05
11	43247172	A	G	rs78850657	9,06E-05
6	111234895	C	T	rs9386956	9,07E-05
12	115845759	T	C	rs1874895	9,07E-05
6	19611159	T	C	rs6934188	9,08E-05
4	174529030	G	A	rs142967203	9,09E-05
4	190065740	A	G	rs13141358	9,11E-05
10	33059585	G	A	rs7915385	9,15E-05
19	42169932	G	C	rs11878851	9,16E-05
19	42169949	A	C	rs11878792	9,16E-05
8	136941472	C	A	rs962158	9,18E-05
16	69036127	A	T	rs1593138	9,18E-05
17	55005863	G	A	rs8064551	9,19E-05
6	5823011	G	T	rs9392716	9,20E-05
10	123831770	C	G	rs11200379	9,21E-05
19	42171906	C	T	rs11083635	9,21E-05

2	69152401	T	C	rs28655947	9,23E-05
10	106994756	A	G	rs1484248	9,24E-05
5	109526971	C	A	rs10042617	9,26E-05
1	64813160	C	T	rs10157897	9,27E-05
5	13838073	C	T	rs2121125	9,30E-05
18	23966596	T	C	rs12962768	9,31E-05
3	65533299	T	C	rs74861858	9,33E-05
7	6130519	C	T	rs852162	9,34E-05
8	41020487	A	T	rs72641807	9,34E-05
11	59775979	T	C	rs555883	9,35E-05
2	69152418	A	T	rs28421715	9,36E-05
6	19629844	A	G	rs2096373	9,36E-05
19	42159840	G	C	rs11880655	9,36E-05
8	41030425	G	A	rs10958664	9,39E-05
9	30837981	G	T	rs10813367	9,39E-05
20	8036018	A	T	rs12480395	9,41E-05
5	13229226	C	G	rs12522860	9,42E-05
5	17993660	G	A	rs704030	9,42E-05
6	36533204	C	A	rs72852326	9,44E-05
6	5822674	G	A	rs13218954	9,46E-05
14	101353211	C	A	rs6575806	9,49E-05
12	115843573	T	C	rs7975091	9,50E-05
14	35265399	T	C	rs76738597	9,53E-05
10	32791058	T	G	rs7090007	9,54E-05
10	107020790	A	G	rs10884125	9,54E-05
12	53039757	T	G	rs11170215	9,56E-05
6	99490353	A	T	rs9321329	9,57E-05
12	119763563	T	G	rs4767808	9,57E-05
10	32757518	A	T	rs944823	9,58E-05
11	43235224	G	C	rs74696691	9,59E-05
1	64813620	T	C	rs10158593	9,62E-05
1	64813657	C	T	rs11208429	9,63E-05
11	43235744	T	C	rs12225754	9,72E-05

11	43235219	T	C	rs79909711	9,73E-05
11	43235304	G	C	rs74377757	9,75E-05
11	43234585	C	T	rs56408379	9,76E-05
7	151611360	T	G	rs80142634	9,77E-05
5	13230105	G	A	rs12523603	9,79E-05
5	32188351	T	A	rs7715242	9,79E-05
11	26279145	T	C	rs11029441	9,79E-05
11	43233898	C	T	rs11037358	9,80E-05
11	43231833	G	A	rs116837421	9,88E-05
10	32771367	G	T	rs3006726	9,90E-05
7	4113494	C	G	rs7807206	9,93E-05
7	146401751	A	G	rs344453	9,93E-05
11	43229632	C	T	rs11530138	9,94E-05
10	32897917	C	T	rs2992079	9,97E-05
11	43229673	T	C	rs11530139	9,97E-05
6	36396909	T	G	rs1742757	9,98E-05
10	118047671	G	A	rs758575	9,99E-05
6	19607690	C	T	rs10946329	1,00E-04
6	19607696	C	T	rs10946330	1,00E-04
8	136944696	A	G	rs2317534	1,00E-04

Supplementary Table 2: Host genes associated with CUD SNPs identified using SNPnexus GAD.

Host Genes
GC
KCTD20
DHFRP2
SEMA3D
PHTF2
PKD1L1
C7orf44
CAP2
NRAS

CALN1
CAPN8
SPEN
EPHB2
PDE4DIP
IL12RB2
MTF1
AFF3
TPO
MLL3
AMD1
RBMXP1
EMR2
EMR3
ANG
HP
HPR
CCDC7
MPP7
C7orf10
C10ORF68
ZSWIM5
PTP4A2
MPDZ
TDPX2
MBL2
WDR7
ZNF32
RET
SORCS3
ANK3
CPN1
LIPA

AGK
STK38
BLK
ANO3
NAV2
SLC6A5
GRIN2A
PDHX
RCN1
HSP90AA2
TMEM132C
DAO
TPH2
ZNF107
ZNF92
ZNF138
CDH3
TNP2
SEPT10
AGBL3
GPR109A
ATG7
LRRN1
ZNF717
TACC2
INPP5F
MYEOV
MADD
ZNF408
SF1
RAB11FIP4
PEMT
SLC39A1

FCRL3
ASAP1
NCAM2
TRIM16
COX10
TKT
TMEM9
HTRA3
AREG
FAM23A
MRC1L1
MRC1
IFI27L1
TGM6
LRRN4
FBLN2
FLJ22536
UNC13C
SDCCAG10
CWC27
C7orf63
FHOD3
MTHFS
TAF4B
JPH2
DAB1
DIO1
GPR98
SLAMF1
MAGI1
PTPRN2
CNTNAP5
ROCK2

FAM13A
TEC
SLC26A5
PGLYRP2
CACNA2D1
CDC14C
APIP
DNAH5
MYO1D
MICAL2
WISP2
CDH4
STK4
CNTNAP2
VIT
PIGF
ECH1
SNTG2
IMP5
MYOM1
UNQ1887
EIF2AK1
HTT
RBMS3
COLQ
MACROD2
NISCH
ETV7
MAL2
SFRS3
ACTBP2
TNIK
MYO16

BAZ1A
ABCB1
RAX
SDK1
UGCGL2
TCERG1L
ERCC1
BCAM
TGFBR2

Supplementary Table 3: SNPs passing nominal significance threshold ($1x10^{-4}$) associated with MUD identified using Plink. Association was calculated using logistic regression for the local MUD cohort.

chr	marker	position	a1	p-value
2	rs10200024	241929228	A	3,27E-06
2	rs10205276	241929621	T	3,27E-06
1	rs2031132	227931560	C	5,91E-06
15	rs4522363	61740334	A	7,01E-06
2	rs4973656	242880755	A	8,18E-06
5	rs10941110	33999079	A	8,39E-06
11	rs61911809	90634964	T	1,02E-05
11	rs61911810	90635074	A	1,02E-05
1	rs4915588	61751502	C	1,18E-05
17	rs11078736	8152930	G	1,18E-05
1	rs204076	29190390	T	1,28E-05
2	rs13407708	37677981	C	1,28E-05
2	rs2361746	30342143	T	1,29E-05
2	rs7569764	241930259	A	1,33E-05
14	rs56151914	74700579	T	1,44E-05
2	rs4675837	241925194	A	1,46E-05
19	rs259244	33145097	C	1,54E-05
14	rs7146440	56379818	A	1,60E-05
1	rs169450	29197728	T	1,80E-05
3	rs361063	138404378	C	1,95E-05
6	rs1923074	91373687	C	2,02E-05
2	rs7579021	167879727	G	2,03E-05
2	rs6707299	167886743	A	2,03E-05
2	rs6726686	167887798	C	2,03E-05

2	rs35251817	242881624	C	2,09E-05
14	rs11624130	98427113	C	2,09E-05
14	rs11623203	98377133	C	2,11E-05
14	rs11627475	98377474	A	2,11E-05
14	rs4905629	98378328	A	2,11E-05
14	rs1551923	98378984	T	2,11E-05
14	rs1869740	98380350	C	2,11E-05
6	rs2179177	166806793	C	2,17E-05
6	rs9358492	22141596	A	2,30E-05
12	rs61941687	125336621	T	2,30E-05
6	rs2208223	91371095	A	2,39E-05
14	rs4905631	98400165	T	2,40E-05
3	rs1501886	117262387	A	2,42E-05
3	rs9826026	117262780	T	2,42E-05
14	rs754604	98424776	A	2,46E-05
2	rs7569435	169122145	A	2,47E-05
14	rs4905628	98371743	T	2,47E-05
14	rs7154130	98374055	T	2,47E-05
14	rs7154949	98374184	A	2,47E-05
11	rs61911811	90635217	C	2,49E-05
11	rs61911812	90635274	G	2,49E-05
11	rs11605879	90641247	C	2,49E-05
11	rs35559918	90644978	A	2,49E-05
10	rs4752655	123897130	C	2,58E-05
2	rs4952097	30320597	C	2,61E-05
14	rs7142731	98379612	G	2,62E-05
11	rs34221146	90639191	G	2,64E-05
19	rs259229	33134425	A	2,66E-05
19	rs259216	33138921	G	2,66E-05
19	rs411328	33140178	G	2,66E-05
19	rs259246	33146154	C	2,66E-05
19	rs259249	33148852	A	2,66E-05
19	rs259251	33149768	A	2,66E-05
19	rs259258	33152523	G	2,66E-05
19	rs259260	33153261	A	2,66E-05
19	rs259265	33154930	G	2,66E-05
19	rs166310	33155187	A	2,66E-05
19	rs259267	33156600	T	2,66E-05
19	rs259269	33156940	C	2,66E-05
19	rs259271	33157100	G	2,66E-05
19	rs259273	33158730	T	2,66E-05
19	rs259275	33159693	G	2,66E-05
19	rs259280	33161452	T	2,66E-05
19	rs259282	33163043	T	2,66E-05
8	rs3134077	119834468	T	2,70E-05

14	rs1888599	98401251	A	2,72E-05
14	rs7150996	98401469	C	2,72E-05
13	rs9544027	76267467	T	2,76E-05
2	rs1400637	169136489	A	2,78E-05
5	rs3195676	34008100	T	2,83E-05
5	rs6875926	34012492	C	2,83E-05
1	rs11585945	214409010	T	2,84E-05
1	rs12754343	214410272	G	2,84E-05
14	rs2124308	98370463	G	2,90E-05
14	rs4905627	98371578	A	2,90E-05
6	rs3914899	137635214	C	2,92E-05
14	rs8020655	98385058	T	2,93E-05
14	rs55864276	98386818	A	2,93E-05
14	rs2776595	98386835	A	2,93E-05
1	rs157222	29212279	A	2,94E-05
11	rs1404527	90609164	T	2,95E-05
11	rs28817341	90616229	T	2,95E-05
11	rs34339096	90619160	A	2,95E-05
11	rs17209415	90625415	T	2,95E-05
12	rs10880190	42306621	A	2,97E-05
11	rs7950792	90571909	A	2,97E-05
19	rs259241	33143701	G	3,02E-05
2	rs13392079	169119178	T	3,03E-05
14	rs242093	69481343	G	3,12E-05
1	rs7513993	214304025	G	3,15E-05
14	rs2526932	73081068	A	3,28E-05
4	rs12642403	54361640	T	3,32E-05
1	rs71638510	27449900	A	3,33E-05
1	rs35159463	27465657	A	3,33E-05
2	rs6547993	30344444	A	3,36E-05
14	rs12588117	98397479	G	3,37E-05
14	rs2803956	73079464	A	3,38E-05
14	rs55972526	98420780	T	3,41E-05
8	rs3133585	119837151	G	3,47E-05
19	rs259256	33151802	T	3,50E-05
2	rs62172717	160167926	G	3,56E-05
14	rs8015434	98365042	A	3,58E-05
14	rs752668	98366896	G	3,58E-05
14	rs73384053	98367678	G	3,58E-05
1	rs204064	29196162	T	3,60E-05
5	rs10512640	36226350	A	3,66E-05
5	rs13154462	36227340	A	3,66E-05
2	rs34981380	160178874	C	3,80E-05
16	rs11646663	81206360	A	3,82E-05
11	rs2140754	90567759	T	3,86E-05

8	rs10955900	119804886	A	3,87E-05
3	rs413147	138510383	A	3,88E-05
2	rs9309717	3495078	G	3,93E-05
4	rs9968386	56771826	T	3,96E-05
14	rs1456998	98400746	T	3,97E-05
14	rs11160421	98355454	A	3,99E-05
14	rs11629121	98357247	C	3,99E-05
4	rs6534699	129651071	T	4,05E-05
4	rs6534700	129651261	A	4,05E-05
4	rs6813824	129651527	C	4,05E-05
10	rs9664665	115549188	A	4,08E-05
1	rs7515135	87829407	G	4,10E-05
1	rs12137864	87829433	A	4,10E-05
1	rs12133199	87829836	C	4,10E-05
1	rs113121917	87830523	A	4,10E-05
1	rs666151	109689339	A	4,10E-05
1	rs548894	109691188	C	4,12E-05
1	rs157221	29211931	C	4,14E-05
14	rs2526933	73080440	C	4,16E-05
14	rs2803959	73080618	T	4,16E-05
11	rs2691828	90596836	A	4,17E-05
6	rs9366856	34405715	A	4,18E-05
6	rs9394218	34405721	A	4,18E-05
6	rs9469780	34406801	C	4,18E-05
1	rs12747137	214326379	C	4,18E-05
15	rs208864	62593375	C	4,18E-05
1	rs4655953	87832695	A	4,22E-05
16	rs1051547	19279380	T	4,24E-05
16	rs11343	19279464	T	4,24E-05
1	rs3219112	226557504	C	4,24E-05
1	rs3754376	226561064	A	4,24E-05
1	rs3219119	226556443	T	4,25E-05
2	rs4952099	30342750	C	4,32E-05
2	rs4286323	30346287	A	4,32E-05
2	rs6735528	30346463	A	4,32E-05
2	rs6741070	30351509	C	4,32E-05
11	rs34294459	90630998	T	4,32E-05
19	rs62101878	30784330	A	4,38E-05
19	rs12984933	30789847	G	4,38E-05
11	rs7101758	90594199	G	4,65E-05
11	rs7105994	90595185	G	4,65E-05
1	rs10494969	214381305	A	4,67E-05
14	rs4900378	98363152	A	4,73E-05
14	rs1456979	98379196	A	4,73E-05
14	rs1456978	98379218	A	4,73E-05

1	rs11590319	214343609	A	4,77E-05
1	rs12727510	214356500	T	4,77E-05
1	rs11120265	214358758	T	4,77E-05
3	rs959236	85198437	T	4,79E-05
4	rs1447046	56764593	C	4,84E-05
4	rs1447047	56764824	G	4,84E-05
6	rs9385609	132762317	C	4,86E-05
4	rs11935692	56736699	A	4,93E-05
4	rs13147883	56743276	T	4,93E-05
14	rs56224513	78480666	T	4,93E-05
2	rs13424940	30350364	C	4,95E-05
1	rs1805414	226573364	A	4,96E-05
1	rs2027439	226575749	T	4,96E-05
1	rs12132600	88834889	C	4,96E-05
11	rs10895506	103542418	T	5,21E-05
12	rs10859980	96305958	G	5,28E-05
2	rs10803830	169073633	T	5,29E-05
2	rs10208207	169074972	T	5,29E-05
2	rs13390712	169075370	A	5,29E-05
10	rs11196495	115550137	A	5,29E-05
14	rs78485589	73099398	A	5,33E-05
10	rs10749145	115551895	A	5,35E-05
14	rs9788484	78468342	G	5,39E-05
6	rs9385608	132762095	T	5,41E-05
22	rs56400210	42651536	C	5,42E-05
11	rs7951455	110823269	G	5,43E-05
5	rs11738800	11609774	T	5,46E-05
2	rs6547992	30332590	T	5,50E-05
14	rs61979971	98428814	T	5,50E-05
11	rs262408	36028562	A	5,52E-05
8	rs1948467	119836572	C	5,60E-05
8	rs1908220	119843056	G	5,60E-05
8	rs1493934	119846666	C	5,60E-05
14	rs9646139	98391624	A	5,74E-05
14	rs11621807	98395584	T	5,74E-05
14	rs11622947	98396293	T	5,74E-05
14	rs61978051	98396570	G	5,74E-05
8	rs13267346	119811174	A	5,82E-05
2	rs11682086	160180940	T	5,82E-05
1	rs11161996	87851420	C	5,88E-05
5	rs4397133	19199614	A	5,92E-05
5	rs10058352	19200415	C	5,92E-05
5	rs4412118	19208360	G	5,92E-05
5	rs10069446	19208407	A	5,92E-05
5	rs6870686	19208702	A	5,92E-05

5	rs10472912	19209362	C	5,92E-05
5	rs10472913	19209758	T	5,92E-05
5	rs12055172	19209782	A	5,92E-05
5	rs10051861	19214129	A	5,92E-05
4	rs749781	56735832	A	5,93E-05
5	rs6897586	94930272	G	5,95E-05
2	rs6743004	30377884	T	5,97E-05
2	rs6709171	30353131	T	5,99E-05
4	rs35978987	54358819	G	6,02E-05
4	rs57297394	54359130	A	6,02E-05
4	rs6839922	54359468	G	6,02E-05
4	rs7691892	54360381	G	6,02E-05
3	rs71310877	190407918	G	6,02E-05
5	rs13179201	36237238	T	6,03E-05
12	rs7299156	110440159	C	6,03E-05
12	rs7299483	110442806	C	6,03E-05
13	rs1824480	66410271	A	6,18E-05
4	rs1031453	72427450	G	6,21E-05
2	rs4665052	160116281	G	6,25E-05
2	rs17494441	160122313	G	6,25E-05
2	rs72961918	160122639	C	6,25E-05
2	rs72961919	160122675	A	6,25E-05
2	rs11692684	160123808	A	6,25E-05
2	rs35074746	160125665	C	6,25E-05
5	rs10454863	36238287	A	6,27E-05
1	rs1341336	226596389	A	6,29E-05
1	rs2793378	226597341	A	6,29E-05
1	rs1104893	226598652	A	6,29E-05
1	rs2793377	226600639	T	6,29E-05
1	rs2249844	226601135	T	6,29E-05
1	rs1828446	226601320	G	6,29E-05
1	rs1828445	226601401	A	6,29E-05
1	rs2793654	226601706	T	6,29E-05
1	rs2666427	226601874	G	6,29E-05
1	rs1865221	226601974	G	6,29E-05
1	rs1865220	226602286	T	6,29E-05
1	rs2695235	226602556	A	6,29E-05
1	rs2793655	226602814	T	6,29E-05
1	rs2793376	226602921	A	6,29E-05
1	rs2695236	226603003	T	6,29E-05
1	rs2695237	226603635	T	6,29E-05
1	rs2695238	226604519	G	6,29E-05
1	rs1341333	226604638	A	6,29E-05
1	rs1341334	226605024	A	6,29E-05
1	rs1341335	226605081	C	6,29E-05

1	rs2695239	226605227	T	6,29E-05
1	rs2570368	226605373	T	6,29E-05
1	rs2136875	226606536	G	6,29E-05
1	rs2695240	226607336	T	6,29E-05
1	rs2570370	226607385	T	6,29E-05
1	rs1858548	226607774	A	6,29E-05
1	rs1858549	226607892	G	6,29E-05
1	rs17716491	214387858	G	6,32E-05
1	rs1353717	214388709	G	6,32E-05
1	rs11584668	214392218	A	6,32E-05
2	rs955896	169131721	A	6,38E-05
1	rs2885120	214385366	T	6,41E-05
1	rs3219090	226564691	C	6,43E-05
2	rs1565073	225634693	T	6,43E-05
1	rs747657	226550924	C	6,44E-05
10	rs10510150	127788934	C	6,44E-05
16	rs7189019	24429397	C	6,44E-05
4	rs12511088	54363262	A	6,54E-05
4	rs17082923	54363481	T	6,54E-05
4	rs12511253	54363801	A	6,54E-05
2	rs1044107	160172048	C	6,58E-05
4	rs2271559	54362676	G	6,59E-05
4	rs2271558	54362833	G	6,59E-05
4	rs12507563	54363194	G	6,59E-05
3	rs7651802	179673116	G	6,63E-05
3	rs7651805	179673117	G	6,63E-05
3	rs9843042	179673799	C	6,63E-05
15	rs2412911	44943094	A	6,64E-05
12	rs9805018	129710951	G	6,65E-05
14	rs2803957	73079751	T	6,74E-05
9	rs12338810	20683404	T	6,75E-05
9	rs10811389	20688837	A	6,75E-05
9	rs10757143	20691188	C	6,75E-05
9	rs7849140	105019413	G	6,77E-05
5	rs12186384	36220004	G	6,80E-05
14	rs11621307	98407722	C	6,80E-05
11	rs961926	90584024	C	6,80E-05
2	rs13018950	169122597	T	6,87E-05
9	rs10821101	96022865	G	6,87E-05
5	rs6556070	172769501	T	6,88E-05
4	rs1118923	81806514	T	6,89E-05
14	rs10132349	55699958	A	6,92E-05
1	rs3118390	68840507	T	6,94E-05
14	rs1952884	56380030	T	6,97E-05
1	rs4659118	119223020	C	6,99E-05

1	rs7536520	119224268	A	6,99E-05
1	rs4076000	68837169	C	7,08E-05
1	rs4265468	68837170	C	7,08E-05
7	rs10479970	2349470	C	7,13E-05
2	rs6706895	157452216	C	7,20E-05
14	rs9646177	98391492	A	7,22E-05
14	rs62014557	40516264	G	7,27E-05
14	rs7145088	40516775	C	7,27E-05
12	rs7302516	129714026	A	7,29E-05
12	rs7310830	129715761	A	7,29E-05
2	rs2361745	30336769	A	7,30E-05
2	rs7588777	30371358	T	7,30E-05
3	rs10934052	110335256	C	7,31E-05
8	rs10283357	131525239	G	7,44E-05
1	rs61781104	21257235	A	7,46E-05
1	rs61781106	21258204	A	7,46E-05
2	rs888012	202561571	A	7,48E-05
2	rs759383	202562026	T	7,48E-05
2	rs11677237	202562449	C	7,48E-05
4	rs4835257	146697205	C	7,49E-05
2	rs11904323	169125080	A	7,53E-05
14	rs61980885	75023369	T	7,54E-05
14	rs61980888	75031469	C	7,54E-05
14	rs11622583	75033329	G	7,54E-05
14	rs11624749	75040245	T	7,54E-05
13	rs2764609	33719567	G	7,54E-05
1	rs12123910	21233316	T	7,55E-05
4	rs4128839	164203507	G	7,56E-05
1	rs112916060	8028386	T	7,57E-05
4	rs6814793	89043484	T	7,58E-05
4	rs35155973	89043881	C	7,58E-05
4	rs4485788	89045485	C	7,58E-05
4	rs113877129	89045907	T	7,58E-05
4	rs34855961	89046294	G	7,58E-05
1	rs9333372	165600766	G	7,62E-05
12	rs79253235	72123035	C	7,62E-05
7	rs2402787	125324851	G	7,69E-05
7	rs9986904	125347374	C	7,69E-05
1	rs204069	29194818	A	7,80E-05
5	rs13185969	36222746	C	7,84E-05
2	rs11689463	160130670	C	7,87E-05
5	rs6866194	94919966	G	7,88E-05
5	rs10075022	94923063	G	7,88E-05
5	rs58373522	94927426	C	7,88E-05
1	rs1006055	214376776	G	7,91E-05

7	rs6597448	154413752	G	7,97E-05
14	rs8013528	55714840	T	7,99E-05
2	rs2166963	169000293	A	8,02E-05
2	rs16854798	169001183	G	8,02E-05
4	rs11939704	146685889	A	8,09E-05
9	rs73513289	94574413	C	8,12E-05
12	rs11171733	56456266	A	8,12E-05
14	rs757655	73104995	C	8,13E-05
18	rs11664972	73807637	C	8,18E-05
1	rs2793382	226583417	C	8,22E-05
1	rs2793380	226588977	C	8,22E-05
1	rs2793657	226590972	T	8,22E-05
9	rs7855954	111152035	C	8,22E-05
6	rs6902708	55445723	G	8,27E-05
5	rs10052586	19198566	G	8,28E-05
5	rs11950345	19198927	G	8,28E-05
5	rs4397134	19199789	C	8,28E-05
5	rs6451029	19201678	A	8,28E-05
5	rs9687338	19202067	A	8,28E-05
5	rs2068059	19202599	G	8,28E-05
5	rs35794281	19202993	T	8,28E-05
5	rs10040567	19204482	G	8,28E-05
5	rs7727629	19204636	A	8,28E-05
5	rs7714511	19204894	G	8,28E-05
5	rs6875122	19205362	G	8,28E-05
5	rs4543270	19205623	A	8,28E-05
5	rs4587077	19205647	G	8,28E-05
5	rs10039320	19206475	T	8,28E-05
5	rs4470766	19206652	A	8,28E-05
5	rs10071746	19209872	A	8,28E-05
5	rs6451040	19210770	A	8,28E-05
5	rs6451041	19210848	A	8,28E-05
5	rs6451042	19210892	A	8,28E-05
5	rs10472238	19211507	G	8,28E-05
5	rs10069183	19211787	T	8,28E-05
5	rs4273605	19212596	C	8,28E-05
5	rs6858971	19213064	T	8,28E-05
5	rs59447723	19213680	A	8,28E-05
5	rs10067330	19214069	T	8,28E-05
6	rs1812254	91363475	G	8,28E-05
6	rs2750442	91363972	G	8,28E-05
5	rs6880453	11612275	T	8,28E-05
2	rs2217627	157458983	A	8,29E-05
9	rs7468319	20692664	T	8,30E-05
2	rs12612984	169127120	C	8,31E-05

11	rs1357822	90588154	G	8,33E-05
14	rs763402	37171951	C	8,39E-05
11	rs4757026	20068489	C	8,40E-05
6	rs3016563	162192856	A	8,42E-05
6	rs3016562	162192874	A	8,42E-05
6	rs9321343	132761433	T	8,50E-05
4	rs28882506	26065733	T	8,51E-05
1	rs6676364	230364875	C	8,55E-05
14	rs2803962	73083240	C	8,56E-05
1	rs12086543	90475947	A	8,56E-05
1	rs1223810	214366917	A	8,58E-05
4	rs17082893	54353275	A	8,58E-05
4	rs11133276	54369040	A	8,58E-05
1	rs2601634	215165933	T	8,59E-05
14	rs8009244	55643752	G	8,63E-05
6	rs7773447	34405046	C	8,65E-05
10	rs2421621	93400101	T	8,66E-05
10	rs7897524	93404374	C	8,66E-05
10	rs7900799	93404478	C	8,66E-05
10	rs10786008	93404791	G	8,66E-05
10	rs4933682	93406388	T	8,66E-05
10	rs4933687	93410399	C	8,66E-05
10	rs10881942	93415020	G	8,66E-05
10	rs4933691	93419957	T	8,66E-05
1	rs61824933	228992953	T	8,67E-05
12	rs7309194	39207946	A	8,69E-05
9	rs6477643	111150567	C	8,74E-05
5	rs4242174	172769871	T	8,80E-05
15	rs7165146	44951174	A	8,81E-05
11	rs2691827	90594585	T	8,84E-05
6	rs7749222	151006811	A	8,86E-05
16	rs41189	49280655	T	8,87E-05
1	rs2209559	68840797	A	8,88E-05
10	rs4935669	58613853	A	8,91E-05
10	rs11005558	58617952	G	8,91E-05
10	rs11005559	58618079	C	8,91E-05
15	rs8042904	89295527	G	9,01E-05
15	rs11855547	89295727	G	9,01E-05
4	rs13117307	56751740	T	9,08E-05
20	rs2144936	6118980	C	9,10E-05
2	rs13018829	169122644	C	9,14E-05
2	rs16855212	169125462	G	9,14E-05
2	rs2954942	169129878	C	9,14E-05
2	rs12618383	160117960	C	9,15E-05
2	rs35362233	160120317	C	9,15E-05

2	rs13015480	160130449	G	9,15E-05
9	rs10122604	20678602	A	9,16E-05
9	rs10964655	20679192	A	9,16E-05
9	rs10123948	20685097	T	9,16E-05
9	rs10117816	20685335	G	9,16E-05
9	rs7029956	20686095	T	9,16E-05
9	rs10964658	20686561	G	9,16E-05
2	rs62153949	97844541	G	9,17E-05
1	rs17422436	87834794	A	9,17E-05
4	rs10016055	81807865	A	9,17E-05
1	rs3118386	68839242	T	9,17E-05
1	rs2225019	68844739	T	9,17E-05
1	rs6662428	215167236	A	9,19E-05
7	rs1419708	125321117	T	9,31E-05
8	rs1436001	58694706	T	9,31E-05
14	rs4904900	92876837	G	9,32E-05
2	rs13002388	169109956	G	9,32E-05
2	rs11902894	169115051	G	9,32E-05
2	rs34752169	169118135	G	9,32E-05
2	rs34374282	169118454	A	9,32E-05
2	rs7568957	169121700	T	9,32E-05
2	rs7569089	169121841	T	9,32E-05
1	rs587727	109685273	A	9,36E-05
1	rs3811488	230371883	G	9,43E-05
1	rs12563741	230375523	G	9,43E-05
10	rs10884356	108513688	T	9,48E-05
4	rs6840343	54359680	T	9,48E-05
4	rs6822201	54359743	C	9,48E-05
2	rs205638	74950067	A	9,49E-05
11	rs4944616	86357637	T	9,49E-05
7	rs4721571	2350184	C	9,50E-05
7	rs4721573	2350433	A	9,50E-05
7	rs4442022	2351981	C	9,50E-05
1	rs6677172	226542114	C	9,53E-05
1	rs10915986	226543113	T	9,53E-05
1	rs872850	69136972	G	9,55E-05
12	rs1733488	63068496	A	9,55E-05
6	rs4330563	110120131	T	9,55E-05
6	rs4072342	110120524	A	9,55E-05
1	rs1354724	64208224	G	9,55E-05
22	rs4822660	26229211	A	9,64E-05
22	rs4822661	26229502	T	9,64E-05
6	rs9371467	151002593	A	9,65E-05
7	rs4731260	125317331	G	9,67E-05
7	rs1419709	125331471	C	9,67E-05

2	rs10865026	109545056	G	9,71E-05
14	rs11845284	55683908	C	9,71E-05
14	rs35082180	55695579	T	9,71E-05
21	rs928771	39663760	G	9,72E-05
18	rs35584308	64914659	T	9,75E-05
6	rs9382508	55439506	A	9,76E-05
8	rs2087998	119852446	G	9,78E-05
2	rs7574710	160133417	A	9,79E-05
2	rs11676586	160136638	A	9,79E-05
5	rs12189111	94928729	T	9,79E-05
5	rs73777201	94931527	C	9,79E-05
17	rs12938531	8125541	G	9,79E-05
2	rs7560573	160115292	C	9,80E-05
2	rs56349910	160115743	T	9,80E-05
2	rs12618342	160117889	T	9,80E-05
1	rs12022191	233651728	A	9,80E-05
10	rs2419869	115549796	A	9,83E-05
10	rs72823238	108517631	G	9,84E-05
10	rs11193039	108520843	C	9,84E-05
10	rs1338760	108523324	C	9,84E-05
1	rs9333379	165601566	C	9,86E-05
2	rs205634	74955353	C	9,87E-05
2	rs205633	74955369	G	9,87E-05
12	rs1095571	39182859	T	9,88E-05
12	rs851931	39184935	A	9,88E-05
12	rs826836	39185406	A	9,88E-05
12	rs73089563	39206506	T	9,88E-05
16	rs194416	49283154	C	9,90E-05
4	rs7682199	54343363	G	9,93E-05
1	rs4548408	214314948	G	9,94E-05
4	rs11734681	54353092	A	9,94E-05
3	rs9872789	85344869	C	9,96E-05
3	rs9883528	85379001	C	9,96E-05
3	rs7645550	178968634	T	9,98E-05
1	rs9442146	99169289	G	1,00E-04
1	rs9434377	99170631	A	1,00E-04

Supplementary Table 4: Host genes associated with MUD SNPs identified using SNPnexus GAD.

Host Genes
PARK7
SPEN
EIF4G3
EPHB2

SLC9A1
MTF1
CAPN8
PARP1
RET
LIPA
ANK3
SORCS1
NAV2
GPR109A
TPH2
DAO
STARD13
CNTNAP2
ANG
SPG11
TNP2
PFAS
RAX
UNC13A
TPO
MPP4
TGM6
MLL3
MYO18B
LARGE
LRRN1
TKT
KCNMB3
PEX5L
HTT
HTRA3
TEC
NRAS
ABCG2
GC
CTNND2
AMACR
C1QTNF3
PPP1R2P3
RBMXP1
PLEKHG1
PARK2
DPP6
TDPX2

ROR2
WNK2
MGST3
CPN1
SF1
PPM1H
SLC25A21
LTBP2
PKD1L2
ZNF138
PIGF
EDAR
GPD2
PIK3CB
LNX1
FLJ22536
CALN1
SLC39A1
GALNT2
ZNF32
AGK
MUC19
SLC26A5
TMEM132D
LMO7
ABCB1
SLC24A4
SYT17
CDH3
CLIP4
PTPRN2
CMYA3
STK39
LRRN4
FIP1L1
CAP2
PTP4A2
AGBL3
ZNF92
YPEL5
NTNG1
SEPT10
HSP90AA2
VIT
XIRP2

NCAM2
LSAMP
SLC4A4
FGF2
ZNF827
SEMA3D
NFIA
ZNF107
ANKRD27
TACC2
ADAM12
SLC6A5
KIAA1797
KIAA1324
C4ORF22
MPDZ
INPP5F
DHFRP2
PHTF2
LDLRAD3
ZNF408
ME3
SCARB1
PUS7
FIG4
HMLL1
CADM2
KCNJ15
C7orf44
PKD1L1
CCDC38

Supplementary Table 5: Top 50 GO biological processes associated with MUD host genes obtained via Enrichr.

Term	P-value	Adjusted P-value	Odds Ratio	Z-Score
positive regulation of sodium ion transmembrane transporter activity (GO:2000651)	3,79 E-05	0,10	43,64	444,23
vocal learning (GO:0042297)	5,72 E-04	0,42	53,33	398,23
positive regulation of sodium ion transmembrane transport (GO:1902307)	5,72 E-04	0,36	53,33	398,23

imitative learning (GO:0098596)	5,72 E-04	0,32	53,33	398,23
positive regulation of autophagy of mitochondrion (GO:1903599)	7,97 E-04	0,41	45,71	326,14
positive regulation of ion transmembrane transporter activity (GO:0032414)	1,17 E-05	0,06	27,83	315,89
positive regulation of ion transmembrane transport (GO:0034767)	1,36 E-03	0,58	35,56	234,79
positive regulation of monooxygenase activity (GO:0032770)	3,85 E-04	0,39	20,87	164,07
neuron cell-cell adhesion (GO:0007158)	3,36 E-03	0,86	22,86	130,21
positive regulation of muscle hypertrophy (GO:0014742)	3,36 E-03	0,82	22,86	130,21
positive regulation of transporter activity (GO:0032411)	3,36 E-03	0,78	22,86	130,21
positive regulation of cation channel activity (GO:2001259)	8,55 E-04	0,40	16,00	113,02
negative regulation of response to wounding (GO:1903035)	4,39 E-03	0,93	20,00	108,56
hyaluronan catabolic process (GO:0030214)	4,39 E-03	0,90	20,00	108,56
positive regulation of cardiac muscle hypertrophy (GO:0010613)	4,39 E-03	0,86	20,00	108,56
dendritic spine morphogenesis (GO:0060997)	4,39 E-03	0,83	20,00	108,56
positive regulation of calcium ion transmembrane transporter activity (GO:1901021)	4,96 E-03	0,87	18,82	99,90
positive regulation of dendrite extension (GO:1903861)	4,96 E-03	0,84	18,82	99,90
detection of mechanical stimulus (GO:0050982)	4,96 E-03	0,82	18,82	99,90
insulin secretion (GO:0030073)	5,55 E-03	0,86	17,78	92,33
regulation of dendrite extension (GO:1903859)	5,55 E-03	0,83	17,78	92,33
positive regulation of action potential (GO:0045760)	3,69 E-02	1,00	26,67	87,97
regulation of oxidative stress-induced neuron intrinsic apoptotic signalling pathway (GO:1903376)	3,69 E-02	1,00	26,67	87,97
negative regulation of transforming growth factor beta production (GO:0071635)	3,69 E-02	1,00	26,67	87,97
regulation of transforming growth factor beta2 production (GO:0032909)	3,69 E-02	1,00	26,67	87,97
synaptic transmission, glycinergic (GO:0060012)	3,69 E-02	1,00	26,67	87,97
regulation of skeletal muscle satellite cell proliferation (GO:0014842)	3,69 E-02	1,00	26,67	87,97

positive regulation of cholesterol storage (GO:0010886)	3,69 E-02	1,00	26,67	87,97
mRNA 3'-splice site recognition (GO:0000389)	3,69 E-02	1,00	26,67	87,97
cellular triglyceride homeostasis (GO:0035356)	3,69 E-02	1,00	26,67	87,97
indolalkylamine metabolic process (GO:0006586)	3,69 E-02	1,00	26,67	87,97
regulation of NAD(P)H oxidase activity (GO:0033860)	3,69 E-02	1,00	26,67	87,97
regulation of adiponectin secretion (GO:0070163)	3,69 E-02	1,00	26,67	87,97
ceramide transport (GO:0035627)	3,69 E-02	1,00	26,67	87,97
regulation of anion channel activity (GO:0010359)	3,69 E-02	1,00	26,67	87,97
cellular response to magnesium ion (GO:0071286)	3,69 E-02	1,00	26,67	87,97
negative regulation of delayed rectifier potassium channel activity (GO:1902260)	3,69 E-02	1,00	26,67	87,97
regulation of mesenchymal stem cell differentiation (GO:2000739)	3,69 E-02	1,00	26,67	87,97
insulin secretion involved in cellular response to glucose stimulus (GO:0035773)	3,69 E-02	1,00	26,67	87,97
sterol import (GO:0035376)	3,69 E-02	1,00	26,67	87,97
negative regulation of interleukin-6 secretion (GO:1900165)	3,69 E-02	1,00	26,67	87,97
L-amino acid import (GO:0043092)	3,69 E-02	1,00	26,67	87,97
cholesterol import (GO:0070508)	3,69 E-02	1,00	26,67	87,97
negative regulation of amino acid transport (GO:0051956)	3,69 E-02	1,00	26,67	87,97
vesicle-mediated transport in synapse (GO:0099003)	1,46 E-03	0,57	13,33	87,02
positive regulation of receptor activity (GO:2000273)	1,59 E-03	0,58	12,97	83,63
regulation of sodium ion transmembrane transporter activity (GO:2000649)	1,59 E-03	0,54	12,97	83,63
regulation of cellular pH (GO:0030641)	1,71 E-03	0,55	12,63	80,45
neuromuscular junction development (GO:0007528)	6,84 E-03	1,00	16,00	79,76
hyaluronan metabolic process (GO:0030212)	7,53 E-03	1,00	15,24	74,50

Supplementary Table 6: Top 50 miRNA associated with MUD host genes obtained via Enrichr TargetScan 2017 database.

Term	P-value	Adjusted P-value	Odds Ratio	Z-Score
hsa-miR-184	0,08	1,00	12,31	31,35
hsa-miR-191	0,08	1,00	12,31	31,35
hsa-miR-346	0,03	1,00	7,27	25,29
hsa-miR-150	0,02	1,00	5,33	21,17
hsa-miR-223	0,02	1,00	4,90	18,35
hsa-miR-432	0,06	1,00	5,25	15,14
hsa-miR-508	0,07	1,00	4,78	12,99
hsa-miR-302B	0,07	1,00	4,64	12,38
hsa-miR-513	0,04	1,00	3,84	12,02
hsa-miR-382	0,07	1,00	4,44	11,53
hsa-miR-499	0,08	1,00	4,16	10,30
hsa-miR-518A-2	0,04	1,00	3,00	9,33
hsa-miR-149	0,06	1,00	3,29	9,05
hsa-miR-26A, hsa-miR-26B	0,04	1,00	2,66	8,50
hsa-miR-335	0,10	1,00	3,68	8,36
hsa-miR-25, hsa-miR-32, hsa-miR-92, hsa-miR-363, hsa-miR-367	0,05	1,00	2,57	7,93
hsa-miR-30A-5P, hsa-miR-30C, hsa-miR-30D, hsa-miR-30B, hsa-miR-30E-5P	0,03	1,00	2,20	7,73
hsa-miR-202	0,13	1,00	3,14	6,31
hsa-miR-452	0,16	1,00	2,86	5,32
hsa-miR-374	0,10	1,00	2,25	5,09
hsa-miR-217	0,16	1,00	2,78	5,07
hsa-miR-510	0,25	1,00	3,56	4,99
hsa-miR-377	0,13	1,00	2,40	4,89
hsa-miR-9	0,09	1,00	1,92	4,52
hsa-miR-218	0,11	1,00	1,99	4,43
hsa-miR-96	0,12	1,00	2,11	4,41
hsa-miR-185	0,18	1,00	2,58	4,40
hsa-miR-130A, hsa-miR-301, hsa-miR-130B	0,11	1,00	1,98	4,35
hsa-miR-324-5P	0,27	1,00	3,20	4,20
hsa-miR-383	0,27	1,00	3,20	4,20
hsa-miR-19A, hsa-miR-19B	0,11	1,00	1,85	4,15
hsa-miR-29A, hsa-miR-29B, hsa-miR-29C	0,11	1,00	1,84	4,09
hsa-miR-103, hsa-miR-107	0,16	1,00	2,17	3,97
hsa-miR-507	0,20	1,00	2,44	3,96
hsa-miR-373	0,17	1,00	2,11	3,75
hsa-miR-200A	0,29	1,00	2,96	3,69
hsa-miR-125B, hsa-miR-125A	0,15	1,00	1,94	3,65

hsa-miR-182	0,15	1,00	1,93	3,62
hsa-miR-22	0,18	1,00	2,07	3,58
hsa-miR-135A, hsa-miR-135B	0,16	1,00	1,89	3,45
hsa-miR-9	0,18	1,00	2,03	3,45
hsa-miR-486	0,31	1,00	2,76	3,27
hsa-miR-526C, hsa-miR-518F, hsa-miR-526A	0,31	1,00	2,76	3,27
hsa-miR-302C	0,19	1,00	1,98	3,23
hsa-miR-345	0,31	1,00	2,67	3,09
hsa-miR-200B, hsa-miR-200C, hsa-miR-429	0,17	1,00	1,70	3,00
hsa-miR-488	0,32	1,00	2,58	2,92
hsa-miR-370	0,25	1,00	2,06	2,84
hsa-miR-384	0,33	1,00	2,50	2,76
hsa-miR-512-3P	0,26	1,00	2,03	2,73

Supplementary Table 7: Genes identified exclusively in the MUD cohort dataset via SNPnexus GAD.

Host Genes
PARK7
EIF4G3
SLC9A1
PARP1
SORCS1
STARD13
SPG11
PFAS
UNC13A
MPP4
MYO18B
LARGE
KCNMB3
PEX5L
ABCG2
CTNND2
AMACR
C1QTNF3
PPP1R2P3
PLEKHG1
PARK2
DPP6
ROR2
WNK2
MGST3
PPM1H

SLC25A21
LTBP2
PKD1L2
EDAR
GPD2
PIK3CB
LNX1
GALNT2
MUC19
TMEM132D
LMO7
SLC24A4
SYT17
CLIP4
CMYA3
STK39
FIP1L1
YPEL5
NTNG1
XIRP2
LSAMP
SLC4A4
FGF2
ZNF827
NFIA
ANKRD27
ADAM12
KIAA1797
KIAA1324
C4ORF22
LDLRAD3
ME3

Supplementary Table 8: Top 50 GO biological processes associated with MUD host genes obtained via Enrichr when using the initial MUD dataset (no imputation).

Term	P-value	Adjusted P-value	Odds Ratio	Z-Score
regulation of SNARE complex assembly (GO:0035542)	1,17E-03	1,00E+00	38,31	258,68
response to muscle stretch (GO:0035994)	1,17E-03	1,00E+00	38,31	258,68

vesicle-mediated transport in synapse (GO:0099003)	1,28E-03	1,00E+00	13,98	93,12
regulation of cellular pH (GO:0030641)	1,49E-03	1,00E+00	13,26	86,31
positive regulation of sodium ion transmembrane transporter activity (GO:2000651)	1,77E-03	1,00E+00	31,35	198,60
lysine metabolic process (GO:0006553)	2,12E-03	1,00E+00	28,74	176,92
carbohydrate derivative transport (GO:1901264)	2,12E-03	1,00E+00	28,74	176,92
lysine catabolic process (GO:0006554)	2,12E-03	1,00E+00	28,74	176,92
regulation of intracellular pH (GO:0051453)	2,26E-03	1,00E+00	11,49	70,03
neuron projection morphogenesis (GO:0048812)	2,69E-03	1,00E+00	5,26	31,11
neuron cell-cell adhesion (GO:0007158)	2,90E-03	1,00E+00	24,63	143,92
regulation of fibroblast migration (GO:0010762)	2,90E-03	1,00E+00	24,63	143,92
positive regulation of muscle hypertrophy (GO:0014742)	2,90E-03	1,00E+00	24,63	143,92
positive regulation of nervous system development (GO:0051962)	3,61E-03	1,00E+00	9,76	54,90
negative regulation of response to wounding (GO:1903035)	3,79E-03	1,00E+00	21,55	120,13
hyaluronan catabolic process (GO:0030214)	3,79E-03	1,00E+00	21,55	120,13
positive regulation of cardiac muscle hypertrophy (GO:0010613)	3,79E-03	1,00E+00	21,55	120,13
dendritic spine morphogenesis (GO:0060997)	3,79E-03	1,00E+00	21,55	120,13

positive regulation of dendrite extension (GO:1903861)	4,28E-03	1,00E+00	20,28	110,60
regulation of dendrite extension (GO:1903859)	4,80E-03	1,00E+00	19,16	102,27
aspartate family amino acid catabolic process (GO:0009068)	5,35E-03	1,00E+00	18,15	94,94
hyaluronan metabolic process (GO:0030212)	6,52E-03	1,00E+00	16,42	82,65
regulation of cardiac muscle hypertrophy (GO:0010611)	6,52E-03	1,00E+00	16,42	82,65
regulation of cardiac muscle contraction by calcium ion signalling (GO:0010882)	7,14E-03	1,00E+00	15,67	77,46
positive regulation of endothelial cell proliferation (GO:0001938)	7,25E-03	1,00E+00	7,61	37,48
vitamin transport (GO:0051180)	7,79E-03	1,00E+00	14,99	72,78
positive regulation of ion transmembrane transporter activity (GO:0032414)	7,79E-03	1,00E+00	14,99	72,78
dendritic spine organization (GO:0097061)	9,17E-03	1,00E+00	13,79	64,71
axonogenesis (GO:0007409)	9,90E-03	1,00E+00	3,85	17,76
positive regulation of synapse assembly (GO:0051965)	1,06E-02	1,00E+00	12,77	58,01
regulation of endothelial cell proliferation (GO:0001936)	1,13E-02	1,00E+00	6,47	28,99
carboxylic acid catabolic process (GO:0046395)	1,22E-02	1,00E+00	11,89	52,37
triglyceride homeostasis (GO:0070328)	1,22E-02	1,00E+00	11,89	52,37
dicarboxylic acid transport (GO:0006835)	1,30E-02	1,00E+00	11,49	49,87

synaptic vesicle endocytosis (GO:0048488)	1,30E-02	1,00E+00	11,49	49,87
axon guidance (GO:0007411)	1,38E-02	1,00E+00	4,34	18,56
positive regulation of phospholipase C activity (GO:0010863)	1,39E-02	1,00E+00	11,12	47,57
dendrite morphogenesis (GO:0048813)	1,57E-02	1,00E+00	10,45	43,43
neuron projection development (GO:0031175)	1,66E-02	1,00E+00	4,11	16,82
cellular macromolecular complex assembly (GO:0034622)	1,75E-02	1,00E+00	9,85	39,85
regulation of synapse assembly (GO:0051963)	1,85E-02	1,00E+00	9,58	38,23
positive regulation of neuron projection development (GO:0010976)	1,94E-02	1,00E+00	5,28	20,81
actin polymerization or depolymerization (GO:0008154)	1,95E-02	1,00E+00	9,32	36,71
clathrin-dependent endocytosis (GO:0072583)	2,25E-02	1,00E+00	8,62	32,69
receptor-mediated endocytosis (GO:0006898)	2,44E-02	1,00E+00	3,65	13,55
bicarbonate transport (GO:0015701)	2,47E-02	1,00E+00	8,21	30,39
positive regulation of epithelial cell proliferation (GO:0050679)	2,50E-02	1,00E+00	4,79	17,67
negative regulation of response to external stimulus (GO:0032102)	2,50E-02	1,00E+00	4,79	17,67
amide biosynthetic process (GO:0043604)	2,81E-02	1,00E+00	7,66	27,38
regulation of MAPK cascade (GO:0043408)	3,11E-02	1,00E+00	3,38	11,73

Supplementary Table 9: Top 50 miRNA associated with MUD host genes obtained via Enrichr using the TargetScan 2017 database, this data was obtained when using the initial MUD dataset (no imputation).

Term	P-value	Adjusted P-value	Odds Ratio	Z-Score
hsa-miR-4671-3p	1,95E-05	1,33E-02	2,60	28,15
hsa-miR-4536	2,81E-04	2,74E-02	2,18	17,81
hsa-miR-4497	3,80E-04	2,88E-02	3,20	25,18
hsa-miR-496	6,67E-04	4,14E-02	2,09	15,31
hsa-miR-4317	8,92E-04	3,58E-02	2,05	14,38
hsa-miR-4664-3p	1,19E-03	4,27E-02	3,47	23,38
hsa-miR-875-5p	1,94E-03	5,76E-02	2,06	12,88
hsa-miR-4745-3p	2,49E-03	5,67E-02	2,35	14,09
hsa-miR-1538	2,49E-03	5,49E-02	2,35	14,09
hsa-miR-4474-5p	2,62E-03	5,27E-02	1,96	11,66
hsa-miR-1282	2,76E-03	5,23E-02	3,07	18,07
hsa-miR-190b	3,01E-03	5,41E-02	1,94	11,25
hsa-miR-190	3,01E-03	5,27E-02	1,94	11,25
hsa-miR-103b	3,40E-03	5,40E-02	2,01	11,42
hsa-miR-3669	3,95E-03	5,86E-02	3,51	19,42
hsa-miR-3197	5,29E-03	6,70E-02	2,34	12,25
hsa-miR-4796-5p	5,78E-03	6,37E-02	1,91	9,83
hsa-miR-502-3p	6,54E-03	6,98E-02	1,81	9,09
hsa-miR-501-3p	6,54E-03	6,87E-02	1,81	9,09

hsa-miR-1180	6,99E-03	7,02E-02	2,87	14,23
hsa-miR-3651	7,86E-03	7,26E-02	2,06	9,99
hsa-miR-1204	8,55E-03	7,69E-02	2,30	10,93
hsa-miR-2964a-5p	8,69E-03	7,71E-02	1,79	8,52
hsa-miR-1295	9,87E-03	8,54E-02	2,25	10,39
hsa-miR-1249	9,98E-03	8,52E-02	2,69	12,41
hsa-miR-3676	1,10E-02	8,85E-02	2,47	11,12
hsa-miR-4508	1,18E-02	8,68E-02	2,03	8,99
hsa-miR-4703-3p	1,19E-02	8,55E-02	1,77	7,86
hsa-miR-648	1,22E-02	8,35E-02	1,74	7,64
hsa-miR-4315	1,22E-02	8,27E-02	2,42	10,68
hsa-miR-574-3p	1,29E-02	8,45E-02	2,80	12,19
hsa-miR-4473	1,41E-02	8,73E-02	1,82	7,76
hsa-miR-1247	1,41E-02	8,69E-02	2,24	9,54
hsa-miR-95	1,63E-02	9,61E-02	2,67	10,99
hsa-miR-4732-5p	1,67E-02	9,69E-02	1,75	7,14
hsa-miR-3201	1,78E-02	1,01E-01	1,70	6,84
hsa-miR-4791	1,78E-02	9,98E-02	1,70	6,84
hsa-miR-3131	1,78E-02	9,90E-02	2,42	9,76
hsa-miR-611	1,78E-02	9,82E-02	2,42	9,76
hsa-miR-3186-5p	1,84E-02	1,00E-01	2,61	10,42
hsa-miR-3200-3p	2,10E-02	1,10E-01	1,74	6,71
hsa-miR-3152-3p	2,11E-02	1,08E-01	1,64	6,33
hsa-miR-4781-5p	2,17E-02	1,10E-01	2,09	8,00
hsa-miR-196a	2,24E-02	1,08E-01	1,72	6,55

hsa-miR-196b	2,24E-02	1,08E-01	1,72	6,55
hsa-miR-4787-5p	2,24E-02	1,07E-01	1,99	7,56
hsa-miR-503	2,27E-02	1,07E-01	1,66	6,27
hsa-miR-1181	2,50E-02	1,12E-01	4,79	17,67
hsa-miR-3622a-5p	2,53E-02	1,12E-01	1,70	6,25
hsa-miR-4694-5p	2,62E-02	1,13E-01	1,63	5,93

Supplementary Table 10: Genes identified exclusively in the MUD cohort dataset via SNPnexus GAD when using the initial MUD dataset (no imputation).

Host Genes
BNC2
SLC24A4
BMP8A
NFIA
UNC13A
SLC4A4
DNM3
PPM1H
FGF2
RGS6
CCND2
PDZD2
KIF1A
DMD
KIAA1324
ARHGEF7
PCDH15
EBF3
NRXN3

BMP8B
CSMD1
SLC25A21
SLC7A8
TMEM132D
AMACR
ZNF827
DPYD
CLIP4
EDAR
SLC45A2
SCARB1
C1QTNF3
SLC28A1
NTNG1
SLC9A1
PARP1
OGDH
ANKRD27
TMEM27
CTNND2
MGST3
EIF4G3
PRMT3
COL22A1
SPG11
KCNMB3
IL1RAPL1
MPP4
SYT17
DEFA3
HN1L
PARK2

LARGE
PPP1R2P3
ODZ4
MUC19

Supplementary Table 11: Top 50 GO biological processes associated with MUD exclusive host genes obtained via Enrichr when using the initial MUD dataset (no imputation).

Term	P-value	Adjusted P-value	Odds Ratio	Z-Score
response to muscle stretch (GO:0035994)	2,74 E-04	6,98E-01	79,37	651,06
regulation of SNARE complex assembly (GO:0035542)	2,74 E-04	4,66E-01	79,37	651,06
positive regulation of sodium ion transmembrane transporter activity (GO:2000651)	4,17 E-04	5,32E-01	64,94	505,40
lysine metabolic process (GO:0006553)	4,99 E-04	5,09E-01	59,52	452,54
carbohydrate derivative transport (GO:1901264)	4,99 E-04	4,25E-01	59,52	452,54
lysine catabolic process (GO:0006554)	4,99 E-04	3,64E-01	59,52	452,54
regulation of fibroblast migration (GO:0010762)	6,86 E-04	4,37E-01	51,02	371,68
positive regulation of muscle hypertrophy (GO:0014742)	6,86 E-04	3,89E-01	51,02	371,68
positive regulation of cardiac muscle hypertrophy (GO:0010613)	9,01 E-04	4,60E-01	44,64	313,03
hyaluronan catabolic process (GO:0030214)	9,01 E-04	4,18E-01	44,64	313,03
positive regulation of dendrite extension (GO:1903861)	1,02 E-03	4,33E-01	42,02	289,44
regulation of dendrite extension (GO:1903859)	1,14 E-03	4,49E-01	39,68	268,75
vesicle-mediated transport in synapse (GO:0099003)	1,51 E-04	7,71E-01	28,96	254,77
aspartate family amino acid catabolic process (GO:0009068)	1,28 E-03	4,65E-01	37,59	250,50
positive regulation of vesicle fusion (GO:0031340)	1,67 E-02	1,00E+00	59,52	243,65
succinyl-CoA metabolic process (GO:0006104)	1,67 E-02	1,00E+00	59,52	243,65
equilibrioception (GO:0050957)	1,67 E-02	1,00E+00	59,52	243,65

sterol import (GO:0035376)	1,67 E-02	1,00E+00	59,52	243, 65
regulation of skeletal muscle satellite cell proliferation (GO:0014842)	1,67 E-02	1,00E+00	59,52	243, 65
positive regulation of cholesterol storage (GO:0010886)	1,67 E-02	1,00E+00	59,52	243, 65
negative regulation of interleukin-6 secretion (GO:1900165)	1,67 E-02	1,00E+00	59,52	243, 65
cellular triglyceride homeostasis (GO:0035356)	1,67 E-02	1,00E+00	59,52	243, 65
regulation of adiponectin secretion (GO:0070163)	1,67 E-02	1,00E+00	59,52	243, 65
positive regulation of sodium ion transmembrane transport (GO:1902307)	1,67 E-02	1,00E+00	59,52	243, 65
cholesterol import (GO:0070508)	1,67 E-02	1,00E+00	59,52	243, 65
positive regulation of fibroblast migration (GO:0010763)	1,67 E-02	1,00E+00	59,52	243, 65
regulation of skeletal muscle contraction (GO:0014819)	1,67 E-02	1,00E+00	59,52	243, 65
regulation of cardiac muscle hypertrophy (GO:0010611)	1,56 E-03	5,32E-01	34,01	219, 77
hyaluronan metabolic process (GO:0030212)	1,56 E-03	4,98E-01	34,01	219, 77
regulation of cardiac muscle contraction by calcium ion signalling (GO:0010882)	1,72 E-03	5,15E-01	32,47	206, 75
positive regulation of lamellipodium morphogenesis (GO:2000394)	1,94 E-02	1,00E+00	51,02	201, 04
regulation of monocyte chemotactic protein-1 production (GO:0071637)	1,94 E-02	1,00E+00	51,02	201, 04
positive regulation of calcineurin-NFAT signalling cascade (GO:0070886)	1,94 E-02	1,00E+00	51,02	201, 04
mitochondrial DNA metabolic process (GO:0032042)	1,94 E-02	1,00E+00	51,02	201, 04
regulation of cardiac muscle cell membrane potential (GO:0086036)	1,94 E-02	1,00E+00	51,02	201, 04
regulation of endothelial cell chemotaxis to fibroblast growth factor (GO:2000544)	1,94 E-02	1,00E+00	51,02	201, 04
detection of molecule of bacterial origin (GO:0032490)	1,94 E-02	1,00E+00	51,02	201, 04
regulation of nitrogen compound metabolic process (GO:0051171)	2,22 E-02	1,00E+00	44,64	170, 01
response to epinephrine (GO:0071871)	2,22 E-02	1,00E+00	44,64	170, 01
negative regulation of exocytosis (GO:0045920)	2,22 E-02	1,00E+00	44,64	170, 01
pyrimidine nucleobase metabolic process (GO:0006206)	2,22 E-02	1,00E+00	44,64	170, 01

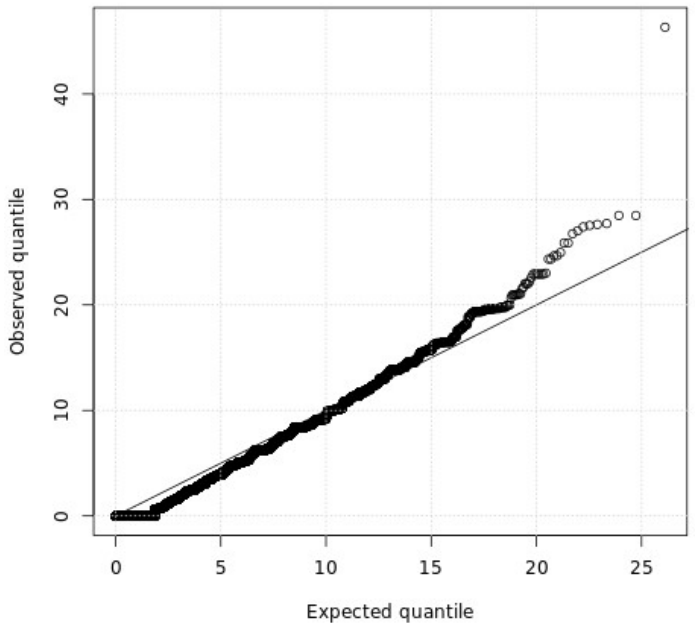
peptidyl-arginine methylation, to asymmetrical-dimethyl arginine (GO:0019919)	2,22 E-02	1,00E+00	44,64	170, 01
lagging strand elongation (GO:0006273)	2,22 E-02	1,00E+00	44,64	170, 01
cellular response to epinephrine stimulus (GO:0071872)	2,22 E-02	1,00E+00	44,64	170, 01
regulation of ATP biosynthetic process (GO:2001169)	2,22 E-02	1,00E+00	44,64	170, 01
polyol biosynthetic process (GO:0046173)	2,22 E-02	1,00E+00	44,64	170, 01
peptidyl-arginine omega-N-methylation (GO:0035247)	2,49 E-02	1,00E+00	39,68	146, 50
peptidyl-arginine methylation (GO:0018216)	2,49 E-02	1,00E+00	39,68	146, 50
membrane disruption in other organism (GO:0051673)	2,49 E-02	1,00E+00	39,68	146, 50
dynamin family protein polymerization involved in mitochondrial fission (GO:0003374)	2,49 E-02	1,00E+00	39,68	146, 50

Supplementary Table 12: Top 50 miRNA associated with MUD exclusive host genes obtained via Enrichr using the TargetScan 2017 database, this data was obtained when using the initial MUD dataset (no imputation).

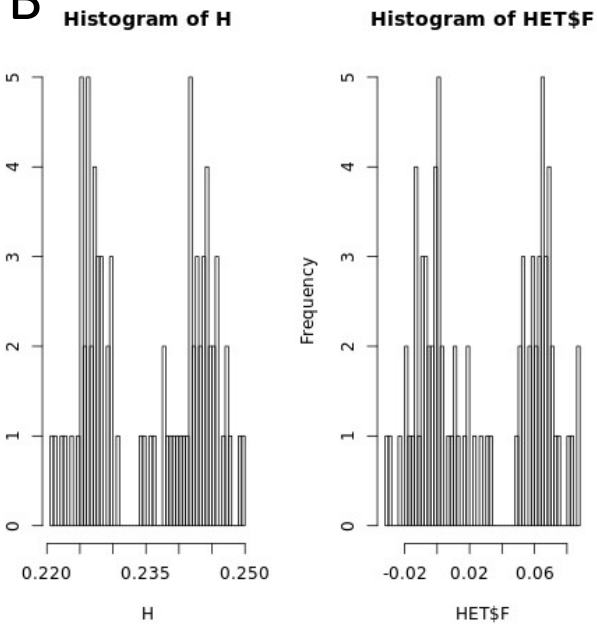
Term	P-value	Adjusted P-value	Odds Ratio	Z-Score
hsa-miR-4536	2,42E-07	1,65E-04	3,73	56,77
hsa-miR-496	9,75E-05	6,66E-03	2,96	27,31
hsa-miR-4497	4,14E-04	1,35E-02	4,42	34,40
hsa-miR-4703-3p	8,86E-04	1,95E-02	2,65	18,66
hsa-miR-3152-3p	9,66E-04	1,69E-02	2,50	17,37
hsa-miR-210	1,17E-03	1,86E-02	4,28	28,89
hsa-miR-3200-3p	1,26E-03	1,87E-02	2,70	18,02
hsa-miR-1247	1,29E-03	1,87E-02	3,71	24,69
hsa-miR-4317	1,50E-03	2,01E-02	2,51	16,30
hsa-miR-503	1,68E-03	2,05E-02	2,48	15,82
hsa-miR-648	1,79E-03	2,04E-02	2,46	15,56
hsa-miR-921	2,10E-03	2,21E-02	2,42	14,89
hsa-miR-3680	2,47E-03	2,51E-02	2,37	14,25
hsa-miR-1282	2,83E-03	2,65E-02	4,23	24,84
hsa-miR-3613-5p	2,89E-03	2,67E-02	2,59	15,12
hsa-miR-4759	2,95E-03	2,62E-02	2,44	14,22
hsa-miR-4671-3p	3,03E-03	2,59E-02	2,57	14,91
hsa-miR-190b	3,28E-03	2,70E-02	2,41	13,78
hsa-miR-190	3,28E-03	2,67E-02	2,41	13,78
hsa-miR-1181	3,45E-03	2,68E-02	9,92	56,25
hsa-miR-2964a-5p	4,05E-03	3,10E-02	2,35	12,94

hsa-miR-4669	4,34E-03	3,26E-02	2,33	12,66
hsa-miR-4796-5p	4,87E-03	3,50E-02	2,42	12,86
hsa-miR-590-5p	5,59E-03	3,90E-02	2,26	11,70
hsa-miR-21	5,59E-03	3,86E-02	2,26	11,70
hsa-miR-219-1-3p	6,10E-03	3,97E-02	2,23	11,38
hsa-miR-1225-5p	6,52E-03	4,20E-02	2,32	11,69
hsa-miR-4520b-3p	6,54E-03	4,13E-02	2,21	11,12
hsa-miR-4684-3p	6,95E-03	4,24E-02	2,19	10,91
hsa-miR-324-5p	7,20E-03	4,28E-02	2,18	10,78
hsa-miR-4745-3p	7,47E-03	4,25E-02	2,78	13,62
hsa-miR-1538	7,47E-03	4,22E-02	2,78	13,62
hsa-miR-4664-3p	8,18E-03	4,47E-02	3,99	19,20
hsa-miR-4473	8,69E-03	4,64E-02	2,36	11,19
hsa-miR-4315	8,77E-03	4,64E-02	3,35	15,86
hsa-miR-1295	8,86E-03	4,62E-02	2,97	14,02
hsa-miR-208a	9,18E-03	4,71E-02	2,22	10,40
hsa-miR-208b	9,18E-03	4,68E-02	2,22	10,40
hsa-miR-3186-5p	9,43E-03	4,77E-02	3,86	17,99
hsa-miR-4529-3p	9,87E-03	4,95E-02	2,19	10,14
hsa-miR-103b	9,92E-03	4,94E-02	2,31	10,66
hsa-miR-4633-5p	1,02E-02	5,04E-02	2,30	10,56
hsa-miR-3117-3p	1,08E-02	5,27E-02	2,28	10,33
hsa-miR-610	1,10E-02	5,32E-02	2,16	9,76
hsa-miR-3923	1,13E-02	5,43E-02	2,41	10,79
hsa-miR-3197	1,14E-02	5,44E-02	2,82	12,64
hsa-miR-4694-5p	1,53E-02	6,63E-02	2,06	8,61
hsa-miR-502-3p	1,54E-02	6,61E-02	2,06	8,60
hsa-miR-501-3p	1,54E-02	6,57E-02	2,06	8,60
hsa-miR-3651	1,57E-02	6,62E-02	2,44	10,14

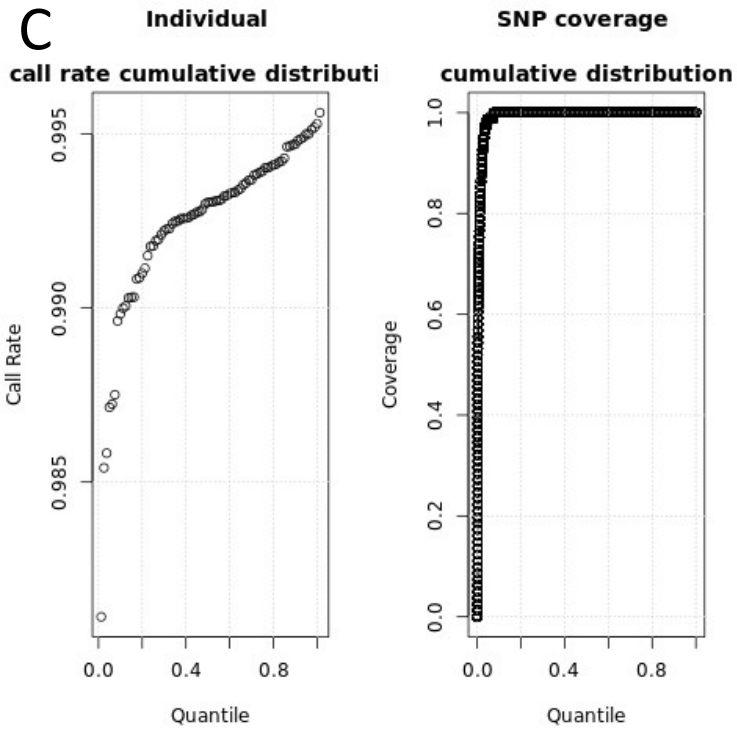
A



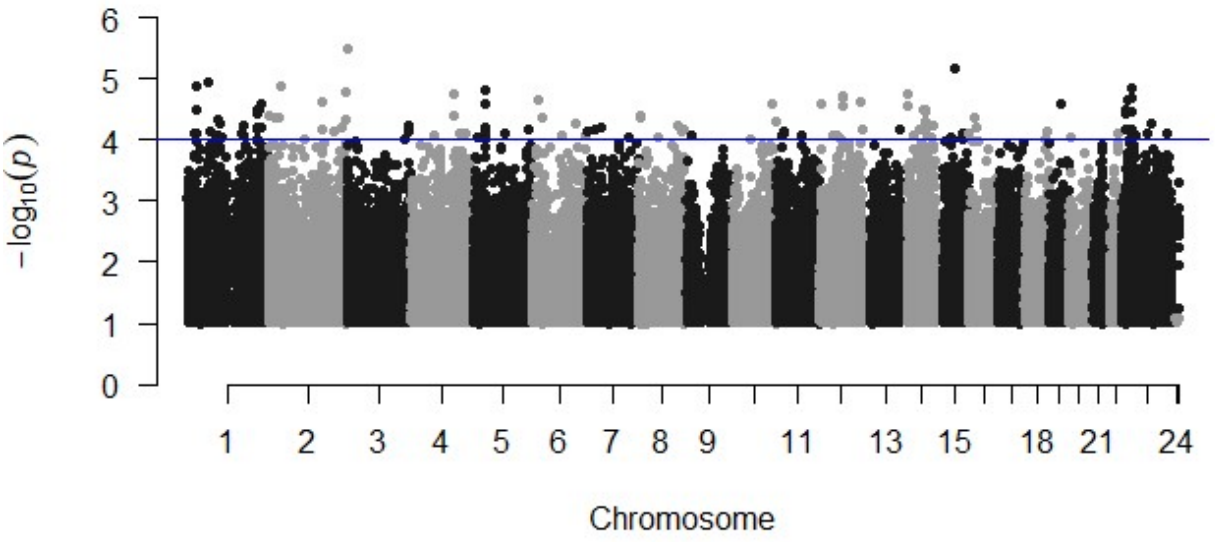
B



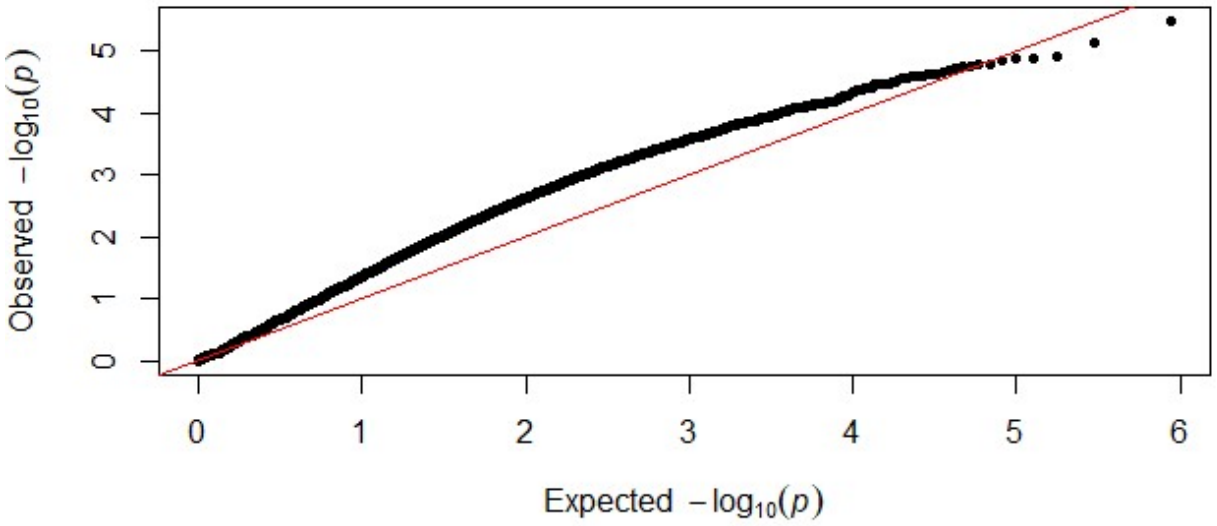
C



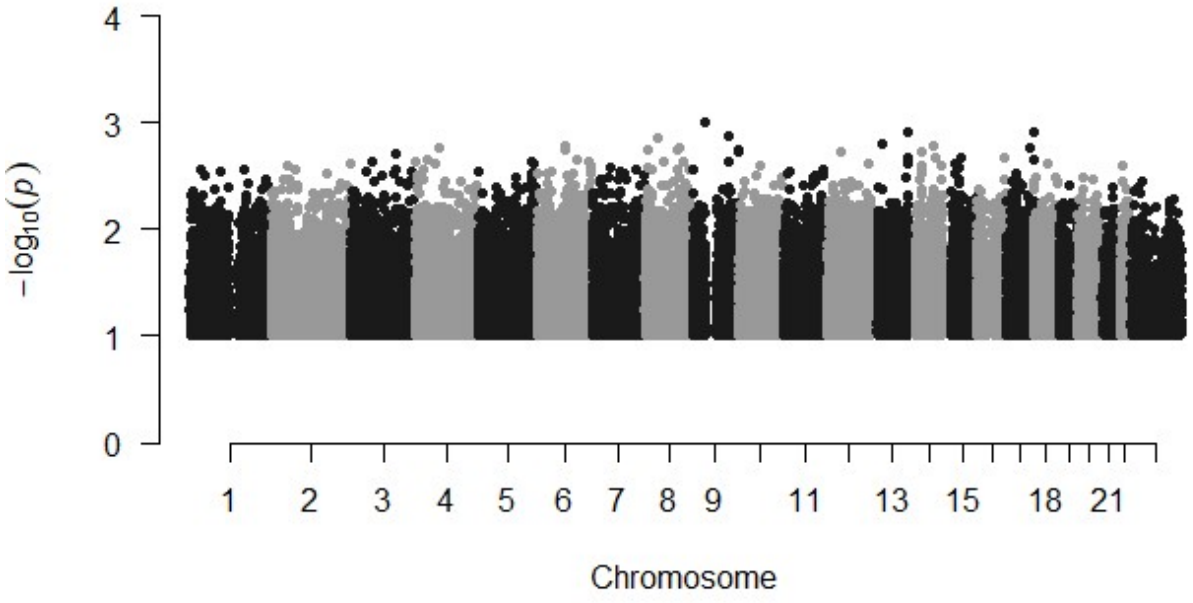
Supplementary Figure 1: A) Q-Q plot of HWE indicating the expected vs observed values. The separation from the expected HWE values indicate SNPs deviating from the applied threshold which were removed following quality control. B) Histograms visualising the heterozygosity-based distributions measured either using F statistics (calculated by plink) or H (calculated using observed homozygotes and the number of non-missing autosomal genotypes). C) Quantile plots of genotype call rates and SNP coverage were used to confirm the viability of the data for further analyses. These indicate the SNPs passing the threshold of 95% (i.e. 95% of individuals in the study were not missing data for aforementioned SNPs), and the coverage of the SNPs read-depths in proportion to the quantiles of the theoretical cumulative distribution function with SNPs presenting low coverage being removed from the data after quality control.



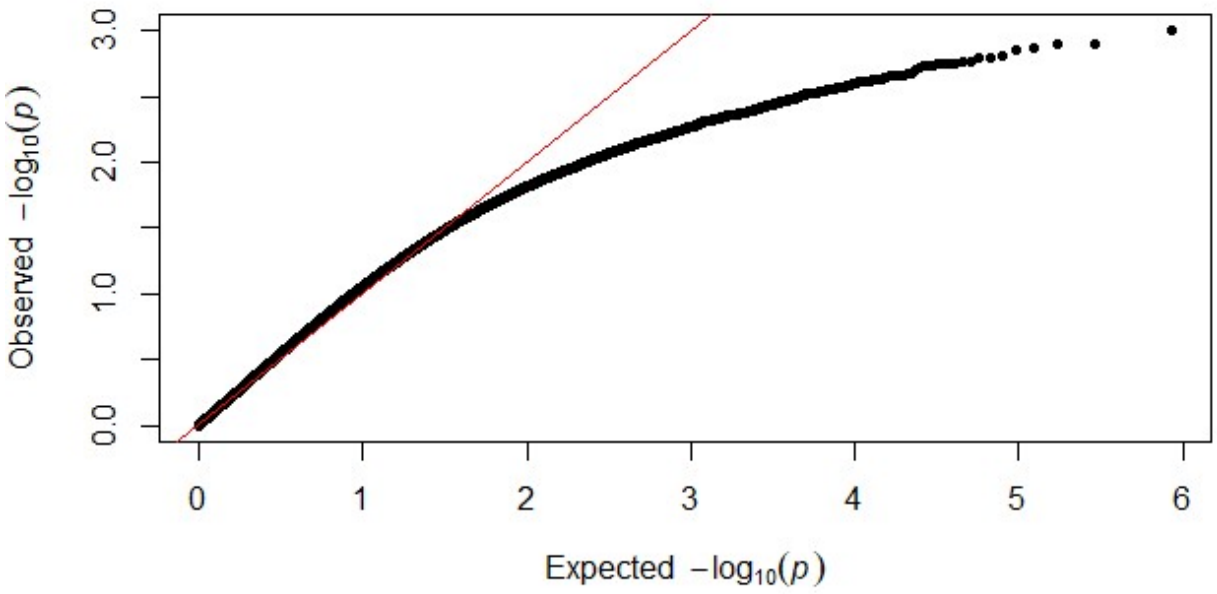
Supplementary Figure 2: Manhattan plot of $-\log_{10}(P)$ logistic regression results for MUD cohort SNPs across each chromosome (where chr 24 refers to the sex chromosomes). A blue 'trend towards significance line' indicates SNPs passing the nominal significance threshold of $1x10^{-4}$ (n = 125).



Supplementary Figure 3: Q-Q plot of $-\log_{10}$(P) results of logistic regression analysis of surviving MUD cohort SNPs plotted against their expected values. Lambda (λ) (genomic inflation) was calculated to be 1.57485 by Plink.



Supplementary Figure 4: Manhattan plot of –log10(P) logistic regression results for MUD cohort SNPs across each chromosome (where chr 24 refers to the sex chromosomes) after adjusting for population stratification using PCs. As expected, once population stratification is adjusted for, the power to detect association decreases and no SNPs pass the nominal (1x10-4) or genome-wide (5x10-8) significance thresholds.



Supplementary Figure 5: Q-Q plot of –log10(P) results of logistic regression analysis of surviving MUD cohort SNPs, after adjusting for population stratification using PCs, plotted against their expected values. Lambda (λ) (genomic inflation) was calculated to be 1.15717 by Plink.